

1984

Probabilistic Causality And The Foundations Of Modern Science

Malcolm Richard Forster

Follow this and additional works at: <https://ir.lib.uwo.ca/digitizedtheses>

Recommended Citation

Forster, Malcolm Richard, "Probabilistic Causality And The Foundations Of Modern Science" (1984). *Digitized Theses*. 1350.
<https://ir.lib.uwo.ca/digitizedtheses/1350>

This Dissertation is brought to you for free and open access by the Digitized Special Collections at Scholarship@Western. It has been accepted for inclusion in Digitized Theses by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlsadmin@uwo.ca.

The author of this thesis has granted The University of Western Ontario a non-exclusive license to reproduce and distribute copies of this thesis to users of Western Libraries. Copyright remains with the author.

Electronic theses and dissertations available in The University of Western Ontario's institutional repository (Scholarship@Western) are solely for the purpose of private study and research. They may not be copied or reproduced, except as permitted by copyright laws, without written authority of the copyright owner. Any commercial use or publication is strictly prohibited.

The original copyright license attesting to these terms and signed by the author of this thesis may be found in the original print version of the thesis, held by Western Libraries.

The thesis approval page signed by the examining committee may also be found in the original print version of the thesis held in Western Libraries.

Please contact Western Libraries for further information:

E-mail: libadmin@uwo.ca

Telephone: (519) 661-2111 Ext. 84796

Web site: <http://www.lib.uwo.ca/>

CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE



National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

PROBABILISTIC CAUSALITY AND
THE FOUNDATIONS OF MODERN SCIENCE

by

Malcolm R. Forster

Department of Philosophy

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
August 1984

6

© Malcolm R. Forster 1984

ABSTRACT

In recent years considerable ink has been spilt in trying to capture the notion of cause in terms of probability. The issue of realism in science has been connected with these attempts notably by van Fraassen, who in the latest book, The Scientific Image, uses Reichenbach's Principle of Common Cause to state the realist's position on scientific research strategies. Van Fraassen then claims, on the basis of his version of Bell's argument, that quantum mechanics is a counterexample to the doctrine of realism.

Van Fraassen's construal of realism is a straw man. A more adequate formulation of realism is developed in the first chapter of this thesis. The essential strategy of science, it is argued, is to aim for a greater unification of its concepts. But it is also difficult to say what exactly unification is, as is well known amongst contemporary philosophers of science. An attempt is made in chapter 1 to capture this notion using the formal ideas of the so-called semantic view of theories as developed by Sneed and others. The result is that unification should be construed as the achievement of simple connections across theory applications, formally construed by Sneed as constraints of classes of models (Sneedian constraints). Less formally the idea is developed as "cross-situational invariance" or "robustness". This less formal treatment connects with Wimsatt's discussion of "robustness as a criterion of reality" as he formulates scientific realism, especially for the biological sciences.

Two crucial difficulties with probabilistic causality are identified in chapter 2. One is the old problem of capturing the asymmetry of 'cause', while the other is a problem of definability when conditional probabilities are 0 or 1. A better theory of causality is developed in chapter 3 under the guidance of the realist principles as explicated in chapter 1. This theory follows the basic ideas of von Wright's "manipulability account of causality", which I refer to as the intervention account (after H. Simon). However, the idea is developed in considerable detail to lay new foundations for the method of path analysis (well known to social scientists).

Finally, van Fraassen's original charge that quantum mechanics is a counterexample to realism is confronted in chapter 5, after formulating some of the conceptual difficulties with quantum statistics in a general way in chapter 4. Admittedly, the realist must give up the

principle of locality, on my view, but it is argued that this need not be a high price for the realist: For it is possible to represent the quantum phenomena in terms of a model of two-way causation that achieves the "cross-situational invariance" desired by the realist without violating the relativistic taboo on causal action along space-like paths. (O. Costa de Beauregard has been pushing a similar view for many years.)

ACKNOWLEDGEMENTS

Special thanks go to my supervisor, Dr. Jeffery Bub, for his help and encouragement, as well as to Drs. William Demopoulos, Bruce Freed, Michael Friedman, William Harper, Cliff Hooker (on leave from the University of Newcastle), and all the staff at the University of Western Ontario. Many of the views presented in chapter 1 are influenced by the work of Professor Cliff Hooker.

Likewise, I owe thanks to my friends Ric Arthur, Balasubramaniam, Jim Brown, John Collier, Pat Enfield, Kristjan Gudmundson, Kai Hahlweg, Ip Po Keung, Si-wai Man, Martin Lockey, Kathleen Okruhlik, Itamar Pitowsky, Edward Manukian, Aldo Mosca, Michael Simpson, and any others I have forgotten, all of whom have contributed to the congenial intellectual atmosphere for studying philosophy of science at Western over the last four and a half years.

Last, but not least, I wish to thank my wife, Jadwiga, as well as my parents, for their support, without which I would not have finished this soon. This thesis is dedicated to my daughter, Natalie, who was born a week before its completion.

TABLE OF CONTENTS

CERTIFICATE OF EXAMINATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
 Chapter 1 - INTRODUCTION	 1
1.1 The 'Cosmic Coincidence' Argument	1
1.2 Other Arguments for Realism	9
1.3 Naturalized Epistemology	16
1.4 Historical Evidence	28
1.4.1 The Copernican Revolution	29
1.4.2 The Acceptance of the Atomic Hypothesis	37
1.4.3 Biological sciences	39
1.4.4 Conclusions	45
1.5 Dispositions, Counterfactuals, and Causality	51
1.5.1 The Problem of Counterfactuals	51
1.5.2 Goodman's Theory of Projectibility	56
1.5.3 'Causality' as a Theoretical Concept ...	69
 Chapter 2 - IN DEFENSE OF THE INTERVENTION THEORY OF CAUSALITY	 76
2.1 Introduction	76
2.2 Two Problems with Probabilistic Causality ...	79
2.2.1 The Problems	79
2.2.2 The solution in Brief	85
2.2.3 Counterfactual Analyses of Causation ...	87
2.2.4 Cartwright's Analysis	91
2.3 Objections to the Intervention Account of Causality	 94
2.4 Path Analysis	100
2.5 Bunzl's Theory of Causal Priority	105
 Chapter 3 - A RE-INTERPRETATION OF PATH ANALYSIS ...	 109
3.1 The Deterministic Case	109
3.2 Generalization to Stochastic Models	124
3.3 What are the Causal Relata? Events or Universals?	 128
3.4 The Question of Transitivity	131
3.5 Epistemological Aspects of the Intervention Account	 133
 Chapter 4 - BELL'S PARADOX AND DETERMINISTIC REALISM	 142
4.1 Introduction	142
4.2 Bell-type Inequalities as Consistency Constraints	 147
4.3 Farkas' Lemma	159

4.4 The Classification of Spin-1/2	
Correlation Functions	169
4.5 Consequences for Realism	174
4.6 Enter Locality!	178
Chapter 5 - BELL'S PARADOX AND STOCHASTIC	
REALISM	190
5.1 Introduction	190
5.2 Why Skyrme's Work is Interesting	193
5.3 The Reduction of Stochastic Realism to	
Deterministic Realism	201
5.4 Some Remarks	208
5.5 A Non-Local Stochastic Model of	
Electron Spin	211
5.6 Final Conclusions	216
BIBLIOGRAPHY	220
VITA	229

Chapter 1

INTRODUCTION

1.1 The 'Cosmic Coincidence' Argument

The purpose of this introductory chapter is two-fold; (1) to make my own views on some aspects of the realism issue explicit so that the reader can better understand the material in later chapters, and (2) to specify and delineate the relevance of the realism issues to this dissertation.

This material will be applied to two topics discussed in later chapters. First, there are the chapters dealing with the notion of causality and its role in scientific understanding, and, secondly, there are two chapters on Bell's theorem construed as an argument against 'hidden variables' as a viable realist strategy for the interpretation of quantum mechanics. Van Fraassen has done me a considerable service in formulating one argument against realism - formulated as the Principle of Common Cause - using the example of quantum mechanical

correlations. This argument will do nicely to lay out roughly the issues under scrutiny in this dissertation.

The only way to understand what the doctrine of realism is meant to assert is to look at the arguments adduced in its favour. Or as Horwich puts it;

The debate surrounding realism is hampered by an aversion to explicit formulation of the doctrine. ... All too often the arguments, for or against, will proceed as though the nature of realism were so well-understood that no careful statement of the position is required. Consequently, several distinct and independent positions have at various times been identified with realism, and the debate is marked by confusion, equivocation and arguments at cross-purposes to one another. [Horwich, 1982, p.181]

The 'cosmic coincidence' argument is commonly cited in favour of realism. Not taking the theoretical 'world' seriously would leave the regularities inherent in our phenomenological experiences a complete mystery, they would be nothing but an unexplained 'cosmic coincidence', and therefore unpalatable. The argument has been applied to argue for common sense realism of the external world. The fact is that our judgements concerning the size of this coffee cup in front of me (readily recognizable on different occasions from shape and colour) does not change when viewed from what are judged to be different distances and different settings. Our judgements of true size, apparent size (the solid angle subtended at our eye), and distance from us conform to certain regular

inter-relationships. The reason for these correlations would remain a complete mystery if it were not for the existence of the external world, the realist contends. Rather than accepting these regularities as cosmic coincidences, we should explain the facts as follows. There is this cup on the table that has intrinsic properties of size, colour, and shape, and, on any given occasion, a set distance from me. The apparent size is a certain function of the true size and the distance from the observer. These physical facts are essential to the explanation of how my judgements are constrained to conform to certain regularities. Moreover, this functional relationship is the same functional relationship that helps explain the correlations inherent in many other diverse instances of our perceptual experience.

The motivation behind Reichenbach's Principle of Common Cause also turns on the 'cosmic coincidence' argument, as van Fraassen describes it;

Reichenbach held it to be a principle of scientific methodology that every statistical correlation (at least every positive dependence) must be explained through common causes. This means then that the very project of science will necessarily lead to the introduction of unobservable structure behind the phenomena. Scientific explanation will be impossible unless there are unobservable entities; but the aim of science is to provide explanation; therefore, the aim of science can only be served if it is true that there are unobservable entities. ... Reichenbach's Principle of Common Cause is that

every relation of positive statistical relevance must be explained by statistical past causes... To put it quite precisely and in Reichenbach's own terms:

If coincidences of two events A and B occur more frequently than would correspond to their independent occurrence, that is, if the events satisfy the relation

$$(1) P(A \& B) > P(A) \cdot P(B),$$

then there exists a common cause C for these events such that the fork ACB is conjunctive, that is, satisfies relations (2)-(5) below:

$$(2) P(A \& B / C) = P(A / C) \cdot P(B / C)$$

$$(3) P(A \& B / \bar{C}) = P(A / \bar{C}) \cdot P(B / \bar{C})$$

$$(4) P(A / C) > P(A / \bar{C})$$

$$(5) P(B / C) > P(B / \bar{C})$$

(1) follows logically from (2)-(5). This principle of common cause is at once precise and persuasive. It may be regarded as a formulation of the conviction that lies behind such arguments as that of Smart, requiring the elimination of 'cosmic coincidence' by science. [van Fraassen, 1980, pp.26-28]

'P' stands for probability and the bar serves to negate the proposition. In opposition, van Fraassen has two objections to this formulation of the 'cosmic coincidence' argument.

I will argue that his principle cannot be a general principle of science at all, and secondly, the postulation of common causes (when it does occur) is also quite intelligible without scientific realism. [van Fraassen, 1980, p.26]

Van Fraassen proceeds to argue these two claims against the principle of common cause. First, that quantum mechanics provides counterexamples to the claim that any correlations can be explained in terms of 'common causes'.

But it is not a principle that guides twentieth-century science, because it is too

close to the demand for deterministic theories of the world that Reichenbach wanted to reject. I will show this by means of a schematic example; but this example will incorporate the sorts of non-classical correlations that distinguish quantum mechanics from classical physics. ... Imagine that you have studied the behavior of a system or object which, after being in state S , always goes into another state which may be characterized by the various attributes F_1, \dots, F_n and G_1, \dots, G_n . Suppose that you have come to the conclusion that this transition is genuinely indeterministic, but you can propose a theory about the transition probabilities: (B) (a) $P(F_i/S) = 1/n$ (b) $P(G_i/S) = 1/n$ (c) $P(F_i \neq G_i/S) = 1$, where \neq means if and only if or when and exactly when. In other words, it is pure chance whether the state to which S transits is characterized by a given one of the F -attributes, and similarly for the G -attributes, but certain that it is characterized by F_i if it is characterized by G_i , by F_j if by G_j , and so on.

If we are convinced that this is an irreducible, indeterministic phenomenon, so that S is a complete description of the initial state, then we have a violation of the principle of common cause. For from (B) we can deduce (9) $P(F_i/S) \cdot P(G_i/S) = 1/n^2$, $P(F_i \& G_i/S) = P(F_i/S) = 1/n$ which numbers are equal only if n is zero or one - the deterministic case. In all other cases, S does not qualify as the common cause of the new state's being F_i or G_i , and if S is complete, nothing else can qualify either. ... Could Reichenbach's principle be weakened so as to preserve its motivating spirit, while eliminating its present unacceptable consequences? ... weakening the principle in various ways (and certainly it will have to be weakened if it is going to be acceptable in any sense) will remove the force of the realist arguments. For any weakening is an agreement to leave some sorts of 'cosmic coincidence' unexplained. But that is to admit the tenability of the nominalist/empiricist point of view, for the demand for explanation ceases to be a scientific 'categorical imperative'. [van Fraassen, 1980, pp.28-31]

I think that van Fraassen's attack here is weak. For

one, it doesn't take any strange quantum mechanical correlations to find a counterexample to Reichenbach's explanatory schema - it is too strong independently of the quantum mechanical example (see chapter 2). We need a different statement of the realist intuition behind the 'cosmic coincidence' argument. But van Fraassen addresses the possibility of finding weaker conditions. "Any weakening", he claims, "is an agreement to leave some sorts of 'cosmic coincidence' unexplained. Perhaps van Fraassen is claiming that any indeterministic theory must leave 'cosmic coincidences' unexplained. But this is absurd. If the data only conform to stochastic regularities, then a stochastic theory can explain all the regularities there. There need not be identified 'cosmic coincidences' left unexplained - and the realist need not believe otherwise. If this were van Fraassen's point then he would not need strange quantum mechanical correlations to make his point. So it is charitable to interpret his argument as merely showing that the Principle of Common Cause fails to tell us how to identify and explain the 'cosmic coincidences' inherent in the quantum mechanical phenomena. The argument only shows what they are not. They are not the kind that can be explained by introducing local hidden variables. The challenge to the realist here is to find a more general principle that does the job.

It is clear that the realist could make his case there

by denying that the state S is complete, but not in the usual sense of postulating hidden variables. The demand for hidden variables is a demand for determinism, and this hinders rather than helps the realist in the case of quantum mechanics [Skyrms (1984) emphasizes this point]. Rather, the realist must claim that S is incomplete in another sense. Namely, that the context dependencies of those correlations must be taken into account. Thus the state S is incomplete in that we need to add propositions about the measurements to be performed on the system before the correlations can be systematically explained. The price paid here is to admit non-locality. The important philosophical issue is then whether non-locality can "conform to the realist motivation". The realist would have to show that such an explanation would not leave any 'cosmic coincidence' unexplained. So, we must ask whether any corrected version of the Principle of Common Cause or a successor to that principle could preserve "its motivating spirit"? We need to identify what this motivating spirit is, and that task will be addressed soon.

So the immediate purpose of this section is to determine what the most general and precise formulation of these realist motivations is. It is only with the completion of this task that any realist interpretation ('nonlocal' or quantum-logical) can be properly

5
evaluated.

Van Fraassen's second objection (that common cause explanations, where they work, can be motivated without scientific realism) is relevant to this more general project:

Nevertheless, there is a problem here that should be faced. Without a doubt, many scientific enterprises can be characterized as searches for common causes to explain correlations. What is the anti-realist to make of this? Is it not a search for explanatory realities behind the phenomena? I think that there are two senses in which a principle of common causes is operative in the scientific enterprise, and both are perfectly intelligible without realism. To the anti-realist, all scientific activity is ultimately aimed at greater knowledge of what is observable. So he can make sense of a search for common causes only if that search aids the acquisition of that sort of knowledge. But it surely does! When past heavy smoking is postulated as a causal factor for cancer, this suggests a further correlation between either cancer and either irritation of the lungs, or the presence of such chemicals as nicotine in the bloodstream, or both. The postulate will be vindicated if such suggested further correlations are indeed found, and will, if so, have aided in the search for larger scale correlations among observable events. [van Fraassen, 1980, p31]

Van Fraassen's general reaction against the 'common cause' formulation of the 'cosmic coincidence' argument is that the search for common causes can be justified solely in terms of empirical adequacy. Where the search is successful, then, we arrive at greater empirical

)

adequacy. But this is a post hoc justification at best. Why should scientists rationally pursue that strategy unless they believe that the search stands some chance of being successful, especially (as van Fraassen points out) that strategy has turned sour in the quantum mechanical domain. Why should scientists continue to believe in that strategy if they didn't believe that the reality underlying the domain of cancer research is different from that of quantum mechanics. 'Promise of empirical adequacy' is not the same thing as 'empirical adequacy'. If van Fraassen wants to introduce the former, then it is he who is making concessions to the realist. For accepting theoretical claims on the basis of believing in the 'promise of future empirical adequacy' is, according to the 'cosmic coincidence' argument, irrational without also believing in the theoretical 'world' described by those theoretical claims. The remaining problem is now the same for van Fraassen and the realists alike: When is it rational to take theoretical claims seriously - whether 'seriously' here means 'believe to be true' or 'believe to promise future empirical adequacy' becomes a rather academic point.

1.2 Other Arguments for Realism

This sort of response to van Fraassen is, I believe, similar to the arguments in defense of realism reported by Friedman (1981,1983) and Earman (1978). Let me first deal with Earman's formulation (which he attributes to F. Ramsey) of the argument.

To make van Fraassen's anti-realist sentiments precise involves formalizing the notion of the empirical content of a theory. As is well known, one way of achieving this is in terms of "Ramseyfication" of a theory - obtained, roughly, by existentially quantifying over all theoretical terms. Let me just explain this in terms of a fictional story. Suppose that there are two theories T_1 and T_2 , each held by different scientific communities. T_1 , imagine, is about the motion of the moon around the earth, and T_2 concerns the motion of the earth around the sun. T_1 has theoretical terms t_1 and t_2 , which represent the masses of the moon and the earth respectively, and T_2 has theoretical terms t_3 and t_4 which are the masses of the earth and sun respectively. The empirical content of T_1 , for instance, is the content of $\text{Ramsey}(T_1)$, which asserts, intuitively, that there exist theoretical functions t_1 and t_2 such that the motion of the moon-earth system accords with the axioms of the theory. $\text{Ramsey}(T_1)$ asserts that there are values that can be assigned as the masses of the moon and the earth such that the theory is consistent with the empirical facts. Similarly for $\text{Ramsey}(T_2)$. The

content of $\text{Ramsey}(T_1)$ and $\text{Ramsey}(T_2)$ can be formalized in terms of the set of all models containing only observational predicates that can be extended to full models of the original theories, T_1 and T_2 . We will denote these model sets as $M(\text{Ramsey}(T_1))$ and $M(\text{Ramsey}(T_2))$. These sets now formally represent the empirical content of the original theories T_1 and T_2 .

Now consider a theory $T_1 \& T_2$, which is about the motion of the moon-earth-sun system. $T_1 \& T_2$, intuitively, contains the laws of motion as they apply to both T_1 and T_2 , as well as the assumption that the mass of the earth is the same with respect to both the component theories T_1 and T_2 , in other words it is assumed that $t_2 = t_3$. But this means that $T_1 \& T_2$ has greater empirical content (it excludes more observational states of affairs) than the empirical contents of T_1 and T_2 combined. $T_1 \& T_2$ might be falsified when neither of T_1 nor T_2 is. In formal terms we have, roughly

$$M(\text{Ramsey}(T_1 \& T_2)) \subset M(\text{Ramsey}(T_1) \& \text{Ramsey}(T_2))$$

$\text{Ramsey}(T_1)$ conjoined with $\text{Ramsey}(T_2)$ fails to include the theoretical claim that the mass of the earth is the same both with respect to moon-earth motion and with respect to earth-sun motion, and is therefore empirically weaker. The process of unification in science involves cross-situational identifications of theoretical

functions,¹ such as $t_2=t_3$ in our example. A strict emphasis on empirical adequacy at each stage of scientific inquiry fails to recognize the importance of such theoretical postulates. Of course, the anti-realist can adopt the full strength of a unified theory such as $T_1 \& T_2$, but only on a post hoc basis.

If the instrumentalist responds by saying that he can take such theoretical claims seriously on the basis of having promise of future empirical adequacy, he is still faced with the same problem as the realist: When do we take these purely theoretical propositions seriously and when not? This is exactly the question that Friedman considers.

Let us ask, then, about the conditions under which a successful theoretical explanation gives us reason to believe in the theoretical 'world' it employs. Here it is necessary to steer a middle course between two extreme views. The first extreme view, which I will call the positivist view, holds that theoretical explanations never give us reasons to believe in the postulated theoretical 'world'. ... It is perfectly possible for the positivistically-minded methodologist to retain theoretical structure - not to try to define it in observational (or other) terms - but, at the same time, to regard such structure as a mere mathematical representation of the empirical facts. On this kind of view, observational or 'phenomenological' facts have the properties they would have if they were embedded in a larger theoretical 'world', but they are not

1. Such postulates are known as Sneedian constraints within the model-theoretic paradigm for theory reconstruction, after their author Sneed (1971).

literally so embedded... [W]e must also be careful to avoid a second extreme view. This view, which has become increasingly popular since the demise of positivism, takes a very liberal and cavalier attitude towards theoretical structure. A theoretical explanation - the embedding of some observational phenomenon in a larger theoretical structure - gives us reason to believe in its postulated structure (or at least provisionally to accept) whenever it is the best available explanation of that phenomenon. ...

I believe that the key to this account is a property of theoretical structure which has often been noticed in the philosophical literature, although not, I think, with the proper emphasis: namely its unifying power. A good or fruitful theoretical structure does not serve simply to provide a model for the particular phenomenon it was designed to explain; rather, in conjunction with other pieces of theoretical structure, it plays a role in the explanation of many other phenomena as well. [Friedman, 1981, pp.2-7]

Re-phrased, the question becomes: Which pieces of theoretical structure are essential to the task of explaining 'cosmic coincidences' and which pieces serve a merely representational function? Again this is the same challenge extracted from van Fraassen's discussion: How to say something more powerful about the deeper realist motives behind the 'cosmic coincidence' argument than that achieved by Reichenbach's Principle of Common Cause? Friedman suggests an answer: The postulated theoretical entity must play a role in the explanation of many different phenomena.

The idea can be pinned down a little more rigorously by making use of some of the ideas underlying the Shoenian

view of theories. Theoretical functions, such as 'mass' or 'weight', can be measured only by making use of the axioms of the theory itself.² Let us take a rather simple mechanical theory as an example. Let D_1 be the set of objects x hanging on a balance at some particular moment. Then our 'theory of the balance' can be stated in terms of the following axioms:

Axiom1: n_1 is a function $n_1: D_1 \rightarrow \mathbb{R}$.

Axiom2: t_1 is a function $t_1: D_1 \rightarrow \mathbb{R}^+$.

Axiom3: $\sum_{x \in D_1} n_1(x) \cdot t_1(x) = 0$.

\mathbb{R} denotes the set of real numbers, and \mathbb{R}^+ the set of positive real numbers. n_1 is a non-theoretical function that assigns positions to each mass on the balance, as measured from the fulcrum. It is measured independently of the theory. t_1 is the theoretical function that, intuitively, assigns the value of its mass to each object x . Its value can only be determined by knowing the function n_1 together with the axioms of the very theory in which the term is embedded. The first two axioms are purely structural, the important axiom scheme being Axiom3. It states the equilibrium conditions for balances, namely that the sum of the force moments about the fulcrum must be zero.

2. A more detailed discussion of Sneed's work appears in Forster, 1978, 1981.

Now consider three distinct applications, 1, 2, and 3, of the theory in which $D_1 = (x_1, x_2)$, $D_2 = (x_2, x_3)$, and $D_3 = (x_3, x_1)$ respectively. Each of these applications will place constraints on the 'mass' values as follows:

Application1: $t_1(x_1) = -[n_1(x_2)/n_1(x_1)] \cdot t_1(x_2)$

Application2: $t_2(x_2) = -[n_2(x_3)/n_2(x_2)] \cdot t_2(x_3)$

Application3: $t_3(x_3) = -[n_3(x_1)/n_3(x_3)] \cdot t_3(x_1)$

The only way that this theory could be falsified is if the balance reached equilibrium with all of the masses on one side of the fulcrum, for this would mean that some masses must be negative, and this is ruled out by axiom2. If the theory is almost tautological in this sense, what sets it apart from other theories? Why take this theory seriously, and therefore the concept of 'mass' implicitly "defined" by that theory, when there are other axiom schemas that might do as well?

The answer is that this particular theory allows us to identify a feature of objects that is invariant across situations. For it turns out to be an empirical fact that $t_3(x_1) = t_1(x_1)$, $t_1(x_2) = t_2(x_2)$, and $t_2(x_3) = t_3(x_3)$. This increases our confidence in the theory to such an extent that we eventually assume the cross-situational invariance of 'mass' and begin to make predictions on that basis. That is, we take 'mass' seriously because it is found.

empirically, to be invariant over a wide variety of situations. In light of this, I conjecture that scientists take a theoretical function seriously exactly when it achieves a certain invariance over a wide range of situations. The realist interprets "taking 'mass' seriously" as "believing in its veridicality". For the instrumentalist, like van Fraassen, it means "promises future empirical adequacy". But even at this stage it is clear that this criterion cuts across the observational/theoretical dichotomy.

1.3 Naturalized Epistemology

It is worthwhile to review well known philosophical arguments against the veridicality of perceptual judgements. Consider again the simple example of estimating the size of an object. Our judgements are invariant over a wide range of circumstances - roughly the same estimates will be given when the cup is viewed on a number of different occasions from different distances and in different settings. But the very existence of perceptual illusions shows that our judgements are not invariant over all circumstances. If we view the cup in an Ames distorted room we may estimate the cup to be bigger than the vase, though that judgement will be

reversed under normal viewing conditions. Why don't we believe that the cup changes its size? Why do we believe that there are such things as illusions?

The main reason, I suggest, is that our attribution of the objective property of size connects with a wide range of scientific explanations, from the thermodynamic properties of objects to their aerodynamical behaviour. Another reason is that the deeper scientific account of the variance of our perceptual judgements only works when we assume that the cup's (real) size is invariant.

The function of the neuronal networks that govern the process of perception is to extract information about the external world from input signals. This process can be viewed as inferential in a generalized sense. The task facing the visual system, for instance, is non-trivial, for the information concerning, say, the 3-dimensional configuration of external objects is underdetermined by the 2-dimensional image that impinges upon our retinas at any particular time. Yet we seem to succeed in making such (unconscious) inferences fairly reliably. The additional information needed to make reliable perceptual judgements must come from the regularities inherent in the sensory data. The essential step in developing a rapport with the external world lies, therefore, in our ability to exploit the deeper regularities hidden in our sensory

inputs over a longer period of time. This is a highly non-trivial computational task.

Thus Ullman has explained the "kinetic depth effect" (the fact that we can accurately judge the 3-dimensional configuration of a number of points from their 2-dimensional projections in a number of different viewpoints) in terms of an ability to exploit the rigidity regularity of everyday objects. Hoffman, 1983, has continued this program using what he calls the planarity regularity and the transversality regularity to explain other inferential abilities of the visual system. In summary, he states,

... vision is an active process whose function is to infer useful descriptions of the world from changing patterns of light falling on the retinas. These descriptions are reliable only to the extent that the inferential processes building them exploit regularities in the visual world such as rigidity, planarity, and transversality.

Perceptual illusions occur when these exploited regularities fail to hold. One such account is summarized by Segal et al. [See also Epstein, 1977, for an interesting collection of articles on visual constancies.]

Such phenomena as size- and shape-constancies and the distorted room phenomena seem to us, as to their originators, obvious examples of perceptual tendencies shaped largely by experience. ... Closest to the view that has shaped the present research is that of Brunswik and the Ames group, for whom perception is "the

process of apprehending probable significances." Allport has succinctly summarized this position as follows: In the process of perceiving an object, the past experience of the organism plays an important part. Basic to the perception process "... is the fact that the organism has built up certain assumptions about the world in which it lives. These assumptions, which are usually unconscious [result in] the attaching of significance to cues". Our general theoretical position can perhaps be epitomized by Brunswik's phrase "ecological cue validity". [Segal et al., 1966, pp.73-4]

1. The so-called optical illusions result, at least in part, from learned habits of inference that possess ecological cue validity.

2. In different physical and cultural environments, different habits of inference are likely to be acquired, reflecting the different ecological validities.

3. For figures constructed of lines meeting in rectangular junctions, there will be a learned tendency among persons dwelling in carpentered environments to rectangularize those junctures, to perceive the figures in perspective, and to interpret them as two-dimensional representations of three-dimensional objects. Such a tendency produces, or at least enhances, the Muller-Lyer illusion and the Sander parallelogram illusion. Since the tendency is assumed to have more ecological validity for people in Western, or carpentered environments, it is predicted that Western peoples will be more susceptible to these illusions than peoples dwelling in uncarpentered environments.

4. The horizontal-vertical illusion results from a tendency to counteract the foreshortening of lines extending into space away from the viewer, so that the vertical in the drawing that is the stimulus for the illusion is interpreted as representing a longer line. Since the tendency has more ecological validity for peoples living mostly outdoors in open, spacious environments, it is predicted that such people will be more susceptible than Western peoples in urban environments. On the other hand, some non-Western groups should be less susceptible to the illusions, e.g. rain forest or canyon dwellers. [Segal et al., 1966, pp.96-7]

In their book, Segal et al. record a great deal of

empirical data that confirms these predictions.

Clearly this explanation presupposes that cups do not actually change their size in non-rectangular rooms! The irregularity of the phenomenal world is explained in terms of the regularities in the outer world. Another way of making the same point is this. Given the broader and deeper understanding of how we make our perceptual judgements; and knowing the way in which the room is distorted, we could correct that judgement. The corrected judgement would be trusted more than our naive judgement exactly because it is invariant over a greater range of situations.

Our belief that there are genuine illusions, that is, false perceptual judgements, is itself a scientific judgement! And scientific judgements, I maintain, are based upon a belief in its postulated entities that achieve a high degree of cross-situational invariance.

Finally, it is worth pointing to the specific content of the above quotation. The unconscious adaptation the regularities of our perceptual system to its environment can be seen as designed (by evolution) to achieve greater cross-situational invariance of our perceptual experience. According to the principle of invariance as a methodological constraint in science, science itself can be viewed as an extension of this adaptive process. [This

is basically the point developed in Hahlweg, 1983.]

The point against the instrumentalist is that he cannot make use of 'invariance' of computed functions (whether these be the functions "computed" by our perceptual systems that lead to perceptual judgements or the theoretical functions computed by scientists) to justify his belief in observational entities, for then he will find that he then has the same, if not more, reason to believe in theoretical entities. If my conjecture turns out to accord with scientific practice, then the instrumentalist has only two options left. He either resorts to some ad hoc justification for reifying observables to the exclusion of theoretical entities, or he can deny their reality as well, in which case he slips back towards an austere idealism. On either count I think the realist has more to offer.

A potential strength the realist position might claim is in the use of the same criterion for observational and theoretical entities alike - the account of "invariance as a criterion of reality" has greater cross-situational invariance itself, and therefore greater reason to take itself seriously (self-justifying the cynic might say). The concept of 'invariance' affords unifying power to the theory in which it is embedded if it enables us to view some aspects of scientific activity as basically similar

to the unconscious inference processes underlying our deeper, usually unconscious, cognitive processes. This idea is not new.

Quine emphasizes this point: the etiology of empirical beliefs generally in humans and other higher organisms ought to be regarded as fundamentally like that of scientific beliefs. This is the fundamental assumption that links the psychological account to the interests of the epistemologist. I will call this assumption the "Quinean link". As Quine points out, it is this assumption that makes the psychological theory something we might call naturalized epistemology. Given this assumption the psychologist is naturally interested in the character of the reasoning involved in social justification of scientific belief. [Stabler, 1982, pp.157-8]

More recently Hahlweg (1983) has developed a new evolutionary epistemology on the idea of viewing "science as an extension of perception". Wimsatt (1980) introduces the idea of 'invariance' as a methodological criterion in science and emphasizes "fruitful connections" that might be had with other areas.

Models can be viewed as tools for detecting phenomena... When the model truly describes the world in relevant respects, the prediction is a veridical perception of the world. On this view, robustness then becomes relative invariance of the thing or phenomenon detected over different at least partially independent modes of detection. This way of putting it points to fruitful connections with concepts and problems in other areas. Thus Donald Campbell has argued that our ordinary notion of an object is of something which is independently detectable or 'diagnosable' in a variety of ways. A physical object is one for which its boundaries, as determined by a variety of different means, roughly coincide. Thus the table on which I write has boundaries which are visually, tactually, auditorily or orally

detectable, and which at least roughly coincide. Ordinary objects in general and the table in particular are thus robust, as I use the term. Something is illusory to the extent to which boundaries derived from one sensory modality are not 'backed up' as expected by correlative boundaries in other sensory modalities. Thus holograms are illusory objects... This association thus points to the use of robustness as a criterion of the reality, veridicality and non-artificiality of a phenomenon or result... [Wimsatt, 1980, p.308]

Of course, it is not the veridicality of the phenomena that is of interest (as Wimsatt suggests in the last line) but the veridicality of the entities postulated to explain the phenomena that the robustness criterion is useful for. In fact, it is exactly the lack of robustness of phenomenological judgements that makes science essential to the argument. "Robustness as a criterion of reality" argues for scientific realism, and against 'common sense' realism.

Hacking does not mention the idea of 'invariance' or 'robustness' but he does emphasize the importance of experimentation in science in virtue of the wide variety of independent "modes of detection" afforded by that experimental practice.

Too often philosophers imagine that microscopes carry conviction because they help us see better. But that is only part of the story. On the contrary, what counts is what we can do to a specimen under a microscope, and what we can see ourselves doing. We stain the specimen, slice it, inject it, irradiate it, fix it. We examine it using different kinds of microscopes that employ optical systems that rely on almost totally unrelated facts about light. Microscopes carry conviction because of

the great array of the interactions and interferences that are possible. When we see something that turns out not to be stable under such play, we call it an artefact and say it is not real,

Likewise, as we move down in scale to the truly un-seeable, it is our power to use these unobservable entities that will make us believe they are there. Yet I blush over the words "see" and "observe". John Dewey would have said that a fascination with seeing-with-the-naked-eye is part of the Spectator Theory of Knowledge that bedeviled philosophy from the earliest times. But I don't think Plato or Locke or anyone before the nineteenth century was as obsessed with the sheer opacity of objects as we have been since. My own obsession with a technology that manipulates objects is, of course a twentieth-century counterpart to positivism and phenomenology. Their proper rebuttal is not a restriction to a narrower domain of reality, namely to what can be positivistically "seen" (with the eye), but an extension to other modes by which people can extend their consciousness. [Hacking, 1982, p.76]

Again the verdict is that "independent detectability", and not observability, decides whether something is real. The "observability" criterion excludes all theoretical entities, while the "independent detectability" criterion does not.

At the heart of the neo-positivist's reification of observables lies the Foundationalist approach to the Theory of Knowledge: Knowledge of the external world (if we have any) is belief that is justified on the basis of observation. This program can be seen as an attempt to break the larger problem of knowledge into two parts;

first we have the problem of perceptual knowledge, and then we have the problem of how we make theoretical inferences on the bases of that sensory information. The presupposition was that we can get on with the latter problem without first having to solve the former problem. The miserable failure of this approach reflects badly on the presupposition on which it was built. So if we stand back and look at the way in which we interact with the world from a scientific viewpoint, we can see that science itself might itself be better understood if we first examine the mechanisms underlying common sense perceptual judgements. Theoretical inference might then be viewed as an extension of the unconscious mechanisms underlying the process of perception. This program of naturalized epistemology has been outlined by Quine and has been discussed by Hahlweg (1983). Quine describes the program in his 1973 Carus lectures:

Berkeley was bent on deriving depth from two-dimensional data for no other reason than the physical fact that the surface of the eye is two-dimensional. But he and the other old epistemologists would have resisted this statement of the matter, because they saw their problem as one of challenging or substantiating our knowledge of the external world. Appeal to physical sense organs in the statement of the problem would seem circular. The building blocks had to be irreducibly mental, and present to consciousness... This fear of circularity is a case of needless timidity, even granted the project of sustaining our knowledge of the external world. The crucial logical point is that the epistemologist is confronting a challenge to natural science that arises within natural science. The challenge runs as

follows. Science itself teaches that there is no clairvoyance; that the only information that can reach our sensory surfaces from external objects must be limited to two-dimensional optical projections and various impacts of air waves on the eardrums and some gaseous reactions in the nasal passages and a few kindred odds and ends. Now, the challenge proceeds, could one hope to find out about the external world from such meager traces? In short, if our science were true, how could we know it? Clearly, in confronting this challenge, the epistemologist may make free use of all scientific theory. His problem is that of finding ways, in keeping with natural science, whereby the human animal can have projected this same science from the sensory information that could reach him according to his science. ... Our liberalized epistemologist ends up as an empirical psychologist, scientifically investigating man's acquisition of science. [Quine, 1973, pp.2-3]

Dretske (1981), too, has broken with foundationalism in a slightly different way. He seeks a more globally unifying view of knowledge, belief, perception, and cognition, by using a concept of 'information' to reach across these interdisciplinary boundaries.

I suggest that the concept of "cross-situational invariance" may have an important future within naturalized epistemology exactly because that concept itself may provide a cross-situational invariance in unifying certain aspects of science, cognition, perception, and even evolution, which the foundationalists tended to draw apart.

I see van Fraassen as championing the old school. Of course, the realist has not yet won the war because the

details of his broader program, embedded within a naturalized epistemology, have not been supplied. Nor will they be provided here. Rather, this thesis undertakes the more modest task of exploring to what extent the idea of "cross-situational invariance" can be seen as a fundamental principle guiding the interpretation and development of scientific theorizing.

1.4 Historical Evidence

The example of 'mass' considered earlier was intentionally over-simplified. If we look at real Newtonian science, we see that the "method of balances" is only one of many theory-dependent ways of measuring 'mass'. We might have used a spring balance, pendulum experiments, buoyancy tests and so on. The empirical fact that all such tests agree is the 'proof' of the 'Newtonian pudding'. The clear demonstration of the unifying power inherent in the Newtonian paradigm lies in the cross-situational invariance of its theoretical entities. On the observational level, the theory is highly chaotic. There is no way of separating the apparent regularities from "cosmic coincidences" lying there crying out to be explained. It is only with the introduction of theoretical concepts, such as 'mass' and 'force' that the cross-situational regularities can even be identified! And if the invariances inherent in the phenomena are successfully identified with the help of a certain postulated entity, it is then that the entity is taken seriously in virtue of this fact. The difficult part of the scientific enterprise is usually in identifying the 'cosmic coincidences', or cross-situational invariants, in

the first place. That is what the mathematization of nature is all about.

1.4.1 The Copernican Revolution

There are numerous historical cases to support this picture. My favourite example is the Copernican revolution. In modern terminology, Copernicus attributes three motions to the earth; (1) a diurnal spin about its poles, (2) an orbital motion about the sun of period one year, and (3) a slow precession of the earth's axis of spin (fixed at the north pole) around the pole of the ecliptic (the perpendicular to the plane in which the planets orbit the sun) to account for the slow "precession of the equinoxes".

In ancient times the motions (1) and (3) were both attributed to the Eighth Sphere of "fixed stars", and since that sphere "drove" the celestial spheres nested inside it, those motions were transferred to all celestial bodies. But there remained an unexplained coincidence in that each planetary model contained an epicycle of exactly one year. By attributing motion (2) to the earth, Copernicus replaced these separate motions with one motion, thereby placing a Sneedian-type constraint among the different planetary models. Immediately, the order of the planets is fixed in Copernicus' system, and the theory

thereby provides independent ways of detecting the motion of the earth, in particular the length of the calendar year. That the empirical data actually fitted this theoretical identification provides a 'cosmic coincidence' type of argument that Copernicus takes full advantage of.

4

In chapter 10 of Book 1 of De Revolutionibus, after effectively criticizing previous attempts to fix the order of the planets, Copernicus admits that he must assume that the distance of the fixed stars is, from us, many orders of magnitude greater than the diameter of the solar system in order to explain the absence of seasonal variations of parallax with the fixed stars. But in spite of the difficulty in swallowing this bitter pill, Copernicus insists that the balance of evidence is on his side.

I think it is easier to believe this than to confuse the issue by assuming a vast number of Spheres, which those who keep Earth at the center must do. We thus rather follow Nature, who producing nothing vain or superfluous often prefers to endow one cause with many effects. Though these views are difficult, contrary to expectation, and certainly unusual, yet in the sequel we shall, God willing, make them abundantly clear at least to mathematicians. [Copernicus, quoted from Kuhn, 1957. p179, my italics]

Copernicus continues on to describe a range of phenomena that fall into place in his system; differences in progression and retrogression of the planets, differences between inferior and superior planets, changes in the luminosity of Mars, and so on. As Copernicus says of his

system, "All these phenomena proceed from the same cause, namely the Earth's motion".

It might be asked why does Copernicus put the earth in motion? Would not the same system, with the sun rotating around the earth and the planets orbiting the sun (Tycho Brahe's later compromise) achieve the same unification? Ravetz (1967) answers that this would sacrifice the regularity of increasing periods of revolution with increasing distance from the sun (cf. Kepler's third law).

It might have occurred to Copernicus in a flash of insight that putting the earth in motion would radically simplify the structure he had erected so far. Or the ancient principle of decreasing speeds with increasing distances may have pointed the way. In this intermediate system the elements of the harmonious sequence of decreasing mean speeds of planets were coupled with a common annual revolution about the sun. Between Venus and Mars lay the earth, with zero speed. This was clearly anomalous. The earth should have a period of revolution between the nine months of Venus and the 2 years of Mars. The choice is obvious: Make it a year. But a period of a year for a revolution of the earth implies a constancy of the sun, and at a stroke the annual motion of the other planets is removed. Harmony is complete. [Ravetz, 1967, pp.97-8]

So the heliocentric device identifies an additional regularity between planetary distances from the sun and the planetary "clocks" (what was later quantitatively described in Kepler's third law).

Kuhn, I think is overly cynical of the force of

Copernicus' argument when he says;

Each argument cites an aspect of the appearances that can be explained by either the Ptolemaic or the Copernican system, and each then proceeds to point out how much more harmonious, coherent, and natural the Copernican explanation is. There are a great many such arguments. The sum of the evidence drawn from harmony is nothing if not impressive.

But it may well be nothing. "Harmony" seems a strange basis on which to argue for the earth's motion, particularly since the harmony is so obscured by the complex multitude of circles that make up the full Copernican system. Copernicus' arguments are not pragmatic. The appeal, if at all, is not to the utilitarian sense of the practicing astronomer but to his aesthetic sense and to that alone. They had no appeal to laymen, who, even when they understood the arguments, were unwilling to substitute minor celestial harmonies for major terrestrial discord. They did not necessarily appeal to astronomers, for the harmonies to which Copernicus' arguments pointed did not enable the astronomer to perform his job better. New harmonies did not increase accuracy or simplicity. Therefore they could and did appeal primarily to that limited and perhaps irrational subgroup of mathematical astronomers whose Neoplatonic ear for mathematical harmonies could not be obstructed by page after page of complex mathematics leading finally to numerical predictions scarcely better than those known before. Fortunately, as we shall discover ... there were a few such astronomers. [Kuhn, 1957, p.181]

Contrary to Kuhn, I find Copernicus' argument impressive. When Kuhn says that the appearances can be explained by either the Ptolemaic or the Copernican system, he is simply missing the point. It is the fact that the new theory succeeds in explaining old regularities and discovering new ones that matters, and we should not

belittle that intellectual achievement.

Kuhn puts great weight on the terrestrial disharmony wrought by Copernicus' system in challenging Aristotelian physics. If the earth were moving, then why does a stone thrown in the air - having no force acting upon it once it leaves the hand - maintain the same motion as the earth?

But we should compare the harmonies versus disharmonies rather carefully. Did Aristotelian physics achieve such tremendous harmony anyway? What's at stake here is not just the terrestrial domain but the unification of the terrestrial and celestial domain as well. The basis of the Aristotelian system was that the natural motion of bodies is towards the center of the universe, viz the center of the earth. That 'explained' terrestrial gravity, but why don't the heavens fall towards the earth? The answer given was that the celestial bodies are constrained by the heavenly spheres and therefore undergo the natural motion under that constraint, namely uniform circular motion. But was this program successful? Were these theoretical devices doing essential work in explaining 'cosmic coincidences'?

Copernicus refers to the breakdown of the old cosmology, and he criticizes the program severely for exactly the things that mark its failure to unify celestial and terrestrial phenomena - the failure of

Ptolemy's equant device to conform with natural uniform circular motion, the failure of the old theories to fix the length of the calendar year and the failure to demonstrate that the concept of relative distance - a concept essential to the terrestrial phenomena - has any robust applicability in the heavenly domain.

The first failure is easily remedied by adding more epicycles and merely served to show that Ptolemy himself had not always taken the Aristotelian program seriously. But Copernicus offers real solutions to the latter two problems. He demonstrates that 'spatial distance' is a robust concept relative to the known celestial phenomena, and identifies interconnections between celestial distances and celestial measures of time. This may seem a small achievement compared to, say, the unification afforded by Newton's universal theory of gravitation, but we must not let hindsight blur our vision. The Aristotelian program had failed on the cosmological front long enough for serious thinkers to give the Copernican idea a chance to guide their thinking.

So Copernicus emphasizes unifying features of his theory on the offensive against the old theory, and also in replying to the anticipated "terrestrial discord" argument. On both fronts he tries to present his theory again as the great unifier. In his reply he anticipates

Galileo's principle of circular inertia (the material in square brackets is a commentary of Kuhn's):

We must admit the possibility of a double motion of objects which fall and rise in the Universe, namely the resultant of rectilinear and circular motion. [This is the analysis advocated earlier by Oresme.] Thus heavy falling objects, being specially earthy, must doubtless retain the nature of the whole to which they belong. ... [Therefore a stone, for example, when removed from the earth will continue to move circularly with the earth and will simultaneously fall rectilinearly towards the earth's surface. Its net motion will be some sort of spiral, like the motion of a bug that crawls straight towards the center of a rotating potter's wheel.]

That the motion of a simple body must be simple is true then primarily of circular motion, and only so long as the simple body rests in its own natural place and state. In that state no motion save circular is possible, for such motion is wholly self-contained and similar to being at rest. But if objects move or are moved from their natural place rectilinear motion supervenes. Now it is inconsistent with the whole order and form of the Universe that it should be outside its own place. Therefore there is no rectilinear motion save of objects out of their place, nor is such motion natural to perfect objects, since [by such a motion] they would be separated from the whole to which they belong and thus would destroy its unity. ... [Copernicus' argument shows how quickly the traditional distinction between the terrestrial and the celestial regions must disappear when the earth becomes a planet, for he is simply applying the traditional arguments about celestial bodies to the earth. Circular motion, whether simple or compound, is the nearest thing to rest. It can be natural to the earth just as it has always been natural to the heavens, because it cannot disrupt the observed unity and regularity of the universe. Linear motion, on the other hand, cannot be natural to any object that has achieved its own place, for the linear motion is disruptive and a natural motion that destroys the universe is absurd.] [Quoted from Kuhn,

1957, 152-3]

While it's true that Copernicus has no concrete or original contribution to make here, he does show that he is sensitive to not merely terrestrial physics, but rather the interconnections between the terrestrial and celestial domains (the earth is so much like the other planets). This is especially so when Copernicus discusses "whether the center of the Universe is or is not the Earth's center of gravity." (Remember Aristotle's account of terrestrial gravity as being natural motion towards the center of the Universe).

Now it seems to me that gravity is but a natural inclination, bestowed on the parts of bodies by the Creator so as to combine the parts in the form of a sphere and thus contribute to their unity and integrity. And we may believe this property to be present even in the Sun, Moon, and the planets, so that thereby they retain their spherical form notwithstanding their various paths. [Copernicus, quoted from Kuhn, 1957, p.154]

The speculation falls short of universal gravitation, but that idea could be seen as a natural further step. It is true that even if Copernicus had made that suggestion, he would still be far short of Newton's achievement. Copernicus only has the expertise to demonstrate unity on the basis of astronomical data. But that is the only realm in which he has much competition with rival theories anyway.

The point is that Copernicus was concerned with

unification of the phenomena, and certainly paid little attention to numerical accuracy. "Harmony" was his big argument - "Nature prefers to endow one cause with many effects", as he put. And given that his work was read and defended, these arguments must have impressed others, especially (as Kuhn points out) as there was little else to recommend his theory.

The resistance to the Copernican theory, I suggest, lay not so much in its discord with Aristotelian physics (although this argument would be used as well), but rather with the disharmony with a broader earth-centered theist cosmology. For the old cosmology simply did not have a great defense to muster.

1.4.2 The Acceptance of the Atomic Hypothesis

Another example where "independent modes of detection" turned out to be crucial was in the acceptance of the atomic hypothesis:

While Clausius was primarily responsible for reviving the kinetic theory and persuading other scientists to take it seriously, I suppose we must criticize him for adjusting the postulates of the theory in somewhat arbitrary ways in order to escape experimental refutation. Indeed his most famous contribution to kinetic theory, the "mean free path" concept, was devised just for this purpose. He had originally assumed that the gas molecules have effectively zero size and simply bounce back and forth between the solid sides of their container. But a Dutch meteorologist, C.H.D. Buys-Ballot, pointed out that such molecules, moving (as Clausius

estimated) at speeds of several hundred meters per second, would diffuse quickly through a large room, contrary to experience. To extricate himself from this difficulty, Clausius (1858) had to attribute to his molecules a diameter large enough so they could not diffuse too rapidly, but small enough to leave unaffected the other deductions from the theory. He had no independent means of estimating the diameter, so we must conclude that this was a purely ad hoc hypothesis as of 1858. ... James Clerk Maxwell read Clausius' paper in English translation in 1859 and figured out a clever way to refute the kinetic theory by deducing from it a falsifiable consequence: the viscosity of a gas of tiny billiard balls must be independent of density and must increase with temperature. ... The year 1865 saw the triumph of the kinetic theory. Maxwell found that the viscosity of air is indeed constant over a wide range of densities. ... The other major event of 1865 was Josef Loschmidt's determination of the size of a molecule, using the mean-free-path formula from the kinetic theory of gases together with the ratio of gas-to-liquid densities. Other estimates giving somewhat similar results (diameter 10^{-8} to 10^{-7} cm), were published soon after by G.J. Stoney, L. Lorenz, and William Thomson. Suddenly the atom was real - for now atoms could be measured, weighed, and counted. But not everyone was convinced, and it was another 40 years before the "real existence of the atom" could be completely established (with the help of statistical mechanics). Even then scientists could not claim to that they could see an atom; by the time that was possible (Mueller 1956), the question of the existence of atoms was a dead issue. [Brush, 1983, pp.266-268]

In 1858 the atomic hypothesis was purely ad hoc in the sense that its parameters were "fudged" - there were no independent measurements of their values. But the idea was clear enough that others were later able to devise ways of undertaking those measurements. This is a case

3

where - unlike the Copernican case - the theory came first and the theoretical measurements came later. Often old experiments can be reinterpreted, but not always. But in 1858, there were no identified 'cosmic coincidences' that the atomic hypothesis could explain where its rivals could not. That changed later, when new measurements were made and after theoretical developments in statistical mechanics were available to interpret old results, then the atomic hypothesis won out. Again, the criterion of observability played no role in regarding the atom as real.

As the Duhem-Quine thesis testifies, it may well be the case that competing theories could make ad hoc adjustments to fit the recalcitrant data, but then the boot is on the other foot. For it is not the aim of theory construction to just "save the phenomena", but to achieve the greatest possible cross-situational invariance or robustness of its theoretical terms in the process. Ad hoc adjustments to the "protective belt" of a theory achieve the former but not the latter goal.

1.4.3 Biological Sciences

3. Although Copernicus did make some crucial observations of his own before being confident in his theory - Ravetz, 1967.

Physical science is not the only source of examples. Ruse (1973) has an interesting reply to the well known charge that the Principle of Natural Selection is nothing but an implicit definition of the concept of 'fitness' and therefore is tautological in the sense of having no empirical content.

What then is the total relationship between fitness and reproductive success? ... [The biologists claim] that given enough organisms a sufficiently high proportion of the fitter are better reproducers, so that the fitter members as a whole have the reproductive edge. Moreover, this claim is indeed analytic, for this is the definition of what is meant by 'fitter'... But it should be realized that it is not on this definition that the biologist rests his empirical claims. His basic claim, other than his claim that a differential reproduction does occur, is one which could well be false, for his claim is that there is a certain constancy about which organisms in a given situation will turn out to survive and reproduce at a rate better than others. It could logically be the case that in a certain situation that genotype carriers A_1A_1 are much more successful at survival and reproduction and hence fitter than genotype carriers A_2A_2 , and in identical types of situation genotype carriers A_2A_2 are much more successful at survival and reproduction and hence fitter than genotype carriers A_1A_1 . If this kind of randomness held, then the biologist's treatment of natural selection would fall to the ground, for he would not then be able to connect up fitness on one occasion with fitness on other occasions. [Ruse, 1973, pp.40-1]

The suggestion here is that the cross-situational connections between fitness values "measured" on different occasions is crucial to giving the Principle of Natural Selection its predictive power. 'Fitness', then, is just

the vehicle in which information about reproductive success is extrapolated from one occasion to another.

I agree that this is the primary function of any theoretical concept. However, to fulfill this function we can only expect to achieve the simple property of cross-situational invariance at the expense of complicating the connections between non-theoretical and theoretical predicates. This is certainly the case with 'mass' in Newtonian mechanics. Cross-situational invariance or constancy of the mass function is only achieved at the expense of having a highly chaotic array of procedures for (theoretically) measuring mass values. The connection between 'mass' and non-theoretical predicates such as 'position' varies considerably from situation to situation. Or take the Copernican case where the mathematics was just as complicated as in previous astronomical theories.

So the question arises; Is evolutionary biology such a lucky domain of inquiry that it manages to achieve cross-situational invariance of its theoretical terms while retaining such a simple connection between 'fitness' and observables such as reproductive success? If not, which way does the tradeoff fall? Towards complicating or replacing the concept of 'fitness' in order to maximize cross-situational invariance or towards retaining a simple

theoretical definition of 'fitness' at the price of losing invariance?

A serious look at the biological sciences will suggest the former' alternative is the case, exactly as the principle of maximizing cross-situational invariance would predict. As Levins puts it:

[The] attempt to integrate fields that have developed independently leads to many difficulties. Some are difficulties in the translation of concepts from one area to another. Darwinian fitness (Wright's W) has to be interpreted in terms of its ecological components, such as intrinsic capacity for increase (Andrewartha and Birch's r_0) and the carrying capacity of the environment for a given genotype (K). Short-term fitnesses of this kind have to be related to the probability of long-term survival on the geological time scale. [Levins, 1968, p.5]

4

So the situation with 'fitness' is not unlike the situation with Newtonian forces. The total Darwinian fitness of a population is broken up into components just as the total force acting on an object - say the coffee cup on the table - is broken into the force of constraint and the force of gravity. This decomposition, I suggest, helps achieve the greatest possible cross-situational invariance for the concept of 'force'.

4. Any modern textbook on evolutionary ecology will explain how the two refined types of fitness values mentioned by Levins arise out of the logistic model of population growth. In fact Ruse now has a chapter on this in his latest book (1983).

For the only way to see any regularities in the acceleration of bodies on different occasions is to classify forces in accordance with various rules about which special forces act, and how strongly, in different situations.⁵

The success of this systematization depends crucially on the hypothesis that forces add vectorially. Only the resultant force is equated to the acceleration of the body. In this way the force of gravity acting on this cup while sitting on the table is the same as when it is falling freely towards the center of the earth, even though the accelerations are different in these situations. Each component force assumes an invariance across situations that the resultant force does not have. We can only conceive of bodies having constant weight, to the extent that they do, by explaining away the cases in which they do not have constant acceleration. This is done by assuming that forces interfere with each other according to the laws of vector addition. And similarly the rate of change of population size (analogous to acceleration) is the (scalar) sum of various selection

5. The systematization of these special force laws is still an active area of research within the mechanics of materials today, and the derivation of governing equations under different forces of constraint and restricted degrees of freedom requires the mathematical power of the Lagrangian framework.

pressures from different sources (analogous to forces). The purpose of this decomposition is the same in both cases.

— If I am right that the key aim of scientific theories is to maximize the cross-situational invariance of its theoretical functions, then it has to be admitted that the biological sciences have not achieved as much as the physical sciences to date. The reason for this may be that natural processes of self-organisation are highly nonlinear, and the required mathematics is poorly understood. Eigen & Schuster (1979) claim that there is an upper limit to the information (in the sense given by the Shannon-Weaver information theory) that can be "carried" by a linear evolutionary system, and so no such system can have the required efficiency in self-replication and "self-repair".

This suggests that the discovery of cross-situational invariants in the biological sciences may require a different kind of mathematics than the linear differential equations that have served the physical sciences so well. (Indeed, the same may be true of some branches of physics that have been forced into the linear mould, such as quantum mechanics - the stochastic model of spin presented in chapter 5, for instance, is nonlinear.) But there is every reason to believe that this is still the aim of

theory construction in the biological and physical sciences' alike.

The point of this digression is to admit an important limitation in what is achieved in chapter 3 of this thesis, for I have not generalized that interpretational schema to the nonlinear case. The assumption of linearity implies an important limitation in scope. But there is no indication that such a generalization will not be forthcoming.

1.4.4 Conclusions

The positivists sought to reify the empirical content of a theory at the expense of taking any of its theoretical structure seriously. In the formalization of what the empirical content of a theory is, Sneed introduced the idea of constraints over a class of models (Sneedian constraints). Most of the constraints identified by Sneed turned out to be simple cross-situational identifications of theoretical functions, but that fact, in itself, was given no significance. I have tried to turn the table on the neo-positivist by claiming that the key desideratum of theory construction is to isolate quantities, such as 'mass', that have the greatest possible cross-situational invariance. 'Weight' for instance has a great degree of

invariance over independent modes of measurement, but that invariance is not complete. Slightly different values are found for different parts of the earth and depending on the position of the moon. So the invariance of 'weight' - to the extent that it holds - is explained in terms of more fundamental cross-situational invariants such as 'mass' (of both the object and the earth). If this picture is correct, then the realist can claim a deep motivation for taking theoretical properties such as 'mass' seriously - the concept itself isolates regularities that mere chance cannot explain.

This principle does not license the liberal, cavalier attitude towards theoretical structure that Friedman warns us of, for not all predicates achieve a high degree of cross-situational invariance. Successful theory construction is no easy intellectual task, and the search for better theories is unending.

This picture helps explain many features of scientific theories that the model-theoretic approach helped expose. Firstly, the mathematical models that apply in different applications of a theory are highly heterogeneous, and the rules by which the equations are derived is a non-trivial part of the theory (I am thinking of the Lagrangian framework for accommodating forces of constraint, as well as the multitude of different constitutive equations used

in continuum mechanics). There is little cross-situational invariance in the equations of motion in Newtonian mechanics.

Rather, the fundamental regularities in physical phenomena are identified in the fundamental physical constants and in the robust properties of matter, all of which are theoretical in origin. Yet, if we didn't take these properties seriously, we might easily trade off their invariance to obtain greater simplicity in the equations governing the phenomena. All we have to do is supply the equation of motion with sufficient "fudge terms", whose values could be determined from the empirical data in each case.

Why don't we condone this as a general research strategy? What would be lost? No less than the predictive power of theory, for the equations will only fit the phenomena "tightly" on a post hoc basis. This is basically the Ramsey-Earman-Friedman line of argument.

The neo-positivists, such as Sneed and van Fraassen, clearly do not want to lose predictive power, but the only rationale they have for not making this tradeoff is "promise of future empirical adequacy". For the theoretical scheme that sacrifices cross-situational invariance still "saves the phenomena" in a minimal sense, and therefore gives empirical adequacy at any given time.

But mere acceptance of theoretical structure for the promise of future empirical adequacy (rather than belief in the theoretical entities) is irrational, the realist contends, for this leaves well identified 'cosmic coincidences' unexplained.

Once the cross-situational invariance of theoretical terms is seen as a key goal in theory construction, some other traditional puzzles in the philosophy of science might also be clarified. The problem of conceptual change arises, roughly, as follows. If the understanding of theoretical terms is a function of the theoretical network of mathematically expressed connections with other quantities, then the meaning of its terms is continually changing as the theory develops or is even replaced. How, then, do we understand theories across "revolutions", whether they be small or large, (for it is an undeniable fact that we do) so in what does the continuity across theory change consist? This is, crudely, the problem of incommensurability.

Wherein lies the continuity in the transition from Kepler's three laws of planetary motion to Newton's theory of gravitation? The attempt to see this as the logical deducibility of Kepler's laws from Newton's fails because only corrected versions of the former follow from the latter [Suppe, 1977, p.173]. The problem, as I see it, is

to isolate those parts of theory change that are relevant to the meaning variance of its terms from those that aren't. Only when we solve this problem can we hope to pinpoint the continuity across theory change. This problem, I conjecture, is the same problem as before - that of steering a middle course in taking theoretical structure seriously, as opposed to regarding it as merely "representational".

The solution suggested here is that we see the retention of cross-situational invariances as providing the essential continuity across theory change. One theory will strictly supercede another when the later theory preserves all the cross-situational invariances accumulated by the earlier one (maybe with some recategorization of the concepts) and succeeds in isolating previously unseen regularities. On the basis of the new extensions of the domains of invariances of the old concepts, new predictions will be tested. Thus Newton's theory ~~retains~~ all the cross-situational invariances of Kepler's laws, to the extent that they are genuine, and identifies many more of its own.

There will often be "Kuhn loss". For instance, Descartes' theory of vortices (or Kepler's use of the idea) explains why all the planets orbit the sun in the same direction. But Newton's theory of universal

gravitation leaves this unexplained. But the loss is still minor relative to the gains.

More detailed case studies are required, but these cannot be provided here. The purpose of the discussion has simply been to suggest that the approach taken here has some prima facie plausibility in cases in which the deductive-nominological model of reduction is known to fail.

1.5 Dispositions, Counterfactuals, and Causality

1.5.1 The Problem of Counterfactuals

The problem of counterfactuals⁷ is to give a universally valid account their understanding - what inference patterns they conform to (their logic) and what we understand their truth conditions to be (their semantics⁸). Neither front has been entirely successful. Of course, we have strong intuitions about what inferences are valid and what the truth conditions are in particular cases, but there don't seem to be rules that apply to every case. Goodman sees this as a serious failure in the philosophy of science.

The analysis of counterfactual conditionals is no fussy little grammatical exercise. Indeed, if we lack the means for interpreting counterfactual conditionals, we can hardly claim to have any adequate philosophy of science. A satisfactory definition of scientific law, a satisfactory theory of confirmation or of disposition terms (and this includes not only

7. For one formulation of the problem, see Goodman, 1965. The volume edited by Harper et al., 1981, records the historical to recent developments in the logic of counterfactuals.

8. I am referring to truth-conditional semantics. for a development of the alternative "conceptual role" semantics see, for instance, Field, 1980. Harper et al., 1981, contains papers on the semantics of conditionals that rather blur the distinction.

predicates ending in "ible" and "able" but almost every objective predicate, such as "red"), would solve a large part of the problem of counterfactuals. Conversely, a solution to the problem of counterfactuals would give us the answer to critical questions about law, confirmation, and the meaning of potentiality. [Goodman, 1965, p.3]

Any advocate of "possible world semantics" knows the problems well. Consider, for instance, the Ramsey-test paradigm for counterfactuals as a guiding principle for their truth evaluation.

First, hypothetically make the minimal revision of your stock of beliefs required to assume the antecedent. Then evaluate the acceptability of the consequent on the basis of this revised body of beliefs. [Harper, 1981, p.5]

Or in the language of possible world semantics, we "go" to the nearest world in which the antecedent is true, and "look" to see whether the consequent is true in that "world". If yes, then the conditional is true, if no, the conditional is false. The problem is to say what the "nearest" world is. Or, in other terms, which actual conditions are "fixed" and which aren't for the purpose of evaluating the truth of the conditional. (Or to say something about the logic of counterfactuals we need to put constraints on the properties of the "nearest world relation".)

Consider a simple example; "If the match had been struck, then it would have lighted". The first

observation is that such conditionals are not truth functional. For in reference to a normal match this conditional is true, while in another instance where I refer to a wet match the same statement is false. In each case both the antecedent and the consequent are false and so the truth value of the conditional cannot be a function of those truth values. So what is it a function of? We imagine the match being struck and ask ourselves "Does it light?" Well, that depends on whether there is enough oxygen, not too much moisture, the chemical composition of the match head, and so on. So it seems that in believing such a simple conditional we have tacit beliefs about a whole range of things some of us have never even heard of!

Van Fraassen gives another nice example of the basic difficulty.

Let us suppose that I say to myself, sotto voce, that a certain fuse leads to a barrel of gunpowder, and then say out loud, "If Tom lit that fuse there would be an explosion". Suppose that before I came in, you had observed to yourself that Tom is very cautious, and would not light any fuse before disconnecting it, and said out loud, "If Tom lit the fuse, there would be no explosion". Have we contradicted each other?

The Ramsey-test principle says that the understanding of a counterfactual conditional is a function of how we would change our beliefs to accommodate the antecedent. If this is an account of our understanding of counterfactuals

(I can't imagine what else it is meant to be) then our understanding of this simple example must vary from person to person, because their understanding of the relevant background conditions will differ. So how do we explain the usefulness and effectiveness of counterfactual discourse if our understanding of such modes of speech differs according to each linguistic agent's particular belief network? This way of putting it is reminiscent of the earlier problem of incommensurability across scientific paradigms, except now we are dealing with incommensurability on a smaller scale, between the conceptual frameworks of different speakers.

But for now the problem at hand is simpler than this: We could consider ourselves quite happy to find universal rules by which belief change operates, so that we could say how any agent would change his beliefs, whatever they happen to be (we could then explain disagreements as arising from differing background beliefs). So, given a particular conceptual framework, we can apply the Ramsey-test and predict what the agent's truth value (or probability) assignment will be. But even this subproblem is not easy to solve. The rationality constraints on belief change are not obvious, in fact future models of theory acceptance at the social level may show that uniformity at the individual level is detrimental to the stability and adaptability of the social enterprise as a

whole. Rationality constraints on individual belief change may not do the job in explaining the nature of scientific change (qua social phenomenon). If this is so, then the Ramsey-test approach to conditionals, and related approaches, are mistaken.

Goodman's understanding of the problem is differently described. His concern, as he states it, is to give an account of counterfactuals that is illuminating in understanding the nature of the scientific enterprise - the nature and dynamics of scientific laws, the relationship between theory and evidence, and the role of dispositional terms in scientific discourse. After stating as much at the beginning of his book (quoted earlier), he proceeds to connect these issues together. His route is through the analysis of dispositional terms.

Intuitively, we can see the truth of some counterfactuals as being a function of the state of the object referred to. So in the previous case "if struck, it would have lighted" is true of this match because this match has certain properties: it has a dry reactive chemical coating that has an ignition temperature easily produced by being struck. Dispositional predicates generally seem to translate to counterfactuals that conform to this intuition. Thus "is soluble in water" translates to "would dissolve if placed in water" and

"flexible" translates to "would bend under suitable pressure". In fact, almost any "objective" predicate such as "red" so translates, in this case to "would appear red under normal lighting". So in the objective sense of the term, something may be red without appearing red, or appear red without being red. All sensory predicates, such as colours, tastes, and sounds, are ambiguous in this way. (Does a tree that falls deep in the forest make a sound?) On the one hand we have the subjective usage, where "red" means "appears red", while the "objective" meaning is only captured, it seems, by the counterfactual assertion "would appear red under normal conditions". And for the example of size judgements used in section 1, we can translate "has size x" to "would be judged to have size x under normal conditions". To go this route in understanding dispositional terms lands us with the vagueness of counterfactual conditionals; what are the relevant conditions to keep fixed in the "nearest possible world"? What are "normal" conditions? The connection between dispositions and counterfactuals is interesting, but, by itself, it doesn't lead anywhere.

1.5.2 Goodman's Theory of Projectibility

Goodman wants to travel another route - he wants to reduce the "problem of dispositions" to what he argues to be the more fundamental idea of "predicate projection".

We may hereafter abbreviate "bends under suitable pressure" as "flexes" and "fails to bend under suitable pressure" as "fails to flex". Now "flexes" and "fails to flex" are mutually exclusive and together they exhaust the realm of things that are under suitable pressure; but neither apply to anything outside that realm. Thus from the fact that "flexes" does not apply to a thing, we cannot in general infer that "fails to flex" does apply. Within the realm of things under suitable pressure, however, the two predicates not only effect a dichotomy, but coincide exactly with "flexible" and "inflexible". What the dispositional predicates do is, so to speak, to project this dichotomy to a wider or even to the universal class of things; and a predicate like "flexible" may thus be regarded as an expansion or projection of a predicate like "flexes". The problem is to define such predicates solely in terms of manifest predicates. [Goodman, 1965, p.44]

Goodman's purpose here is to analyse dispositional discourse not in terms of the intensional language of counterfactual conditionals (for that just brings us back to the original problem), but to view them as ordinary extensional predicates. (This accords better with the Class Nominalism he subscribes to.) The problem remaining, then, is to characterize the nature of the "projection" from "manifest" predicates, such as "flexes", to dispositional terms such as "flexible". The advantage of revamping the problem this way is to exorcise possibilia from the analysis. Or as Goodman puts it:

My main purpose here, then, has been to suggest that discourse, even about possibles, need not transgress the boundaries of the actual world. What we often mistake for the actual world is one description of it. And what we mistake for possible worlds are just equally

true descriptions in other terms. We have come to think of the actual world as one among many possible worlds. We need to repaint that picture. All possible worlds lie within the actual one... Possible processes and possible entities vanish. Predicates pertaining to them are seen to apply to actual things, but to have extensions related in peculiar ways to, and usually broader than, the extensions of certain manifest predicates. A predicate ostensibly of possibles as compared to a correlative manifest predicate, like an open umbrella as compared to a closed one, simply covers more of the same earthly stuff. [Goodman, 1965, pp.56-7]

Quine expresses a similar attitude towards dispositional terms:

Each disposition, in my view, is a physical state or mechanism. A name for a specific disposition, e.g. solubility in water, deserves its place in the vocabulary of scientific theory as a name of a particular state or mechanism... Where the general dispositional idiom has its use is as follows. By means of it we can refer to a hypothetical state or mechanism that we do not yet understand, or to any of the various such states or mechanisms, while merely specifying one of its characteristic effects, such as dissolution upon immersion in water... The dispositional idiom is indifferent to the physical subject matter and serves only to signal how we are getting at it. So, if I were to devise an ideal language for a finished theory of reality, or any part of it, I would make no place in it for the general dispositional idiom. In developing a theory, on the other hand, the idiom is indispensable. ... And since scientific theory is always developing, the idiom is here to stay. [Quine, 1974, pp.10-12]

Dummett characterizes this way of describing dispositions of a system to respond in a certain way under "counterfactual tests" of suitably restricted kinds as a realist assumption:

The ... assumption, that, for every test (of some suitably restricted kind), there exists a property which is revealed by the test, and which, at any given time, each object either possesses or fails to possess, is not an operationalist assumption as such, but a realist one. The possession or non-possession of a given property P is what gives substance to the truth of counterfactual statements about what the result of the test T, if it had been applied at a given time, would have been; the bivalence assumed for statements of the form "The system S has property P" guarantees that one out of every pair of opposite counterfactuals of this kind must be true. Conversely, the supposition that, of each pair of opposite counterfactuals, there must always be one which is true, is tantamount to the realist assumption: the possession of the property can then be equated to the (hypothetical) satisfaction of the test, and will then, in virtue of the supposition about counterfactuals, satisfy the law of bivalence. [Dummett, 1980, pp.277-78]

Kripke also repudiates the possible world paradigm and the "counterpart" theory of cross-world identity in defending his own theory of rigid designation.

All this talk [of counterparts] seems to have taken the metaphor of possible worlds too seriously in some way... Who is to prevent us from saying "Nixon might have gotten Carswell through had he done certain things"? We are speaking of Nixon and asking what, in certain counterfactual situations, would have been true of him. We can say that if Nixon had done such and such, he would have lost the election to Humphrey. Those I am opposing would argue, "Yes, but how do you find out if the man you are talking about is in fact Nixon"? ... Possible worlds are not something to which an epistemological question like this applies... If we say, "If Nixon had bribed such and such a Senator, Nixon would have gotten Carswell through," what is given in the very description is that it is a situation in which we are speaking of Nixon, and of Carswell, and of such and such a Senator. ... Advocates of the other view take speaking of certain qualities as unobjectionable. They do not say, "How do we know that this quality (in another possible

world) is that of redness?" ... But I see no more reason to object in one case than the other. [Kripke, 1977, pp.80-82]

The move away from possible world semantics taken by all of these authors is a sensible one. Friedman agrees and provides us with a more general argument to that effect.

Now there is no doubt that possible-world semantics has provided a mathematical representation (although essentially a trivial one) for various intensional phenomena. This is just what is shown by the numerous 'completeness' results in 'intensional logic'. But should we also think that it is a genuine theoretical explanation or reduction, with ontological claims that should be taken seriously, say, as the molecular model of a gas? If we think of the postulation of theoretical structure as merely a matter of inference to the best explanation, we will find this conclusion hard to avoid. We have seen above, however, that this cavalier attitude towards theoretical postulation is quite inadequate. The best available explanation must also lead in the right theoretical direction: it must lead to an increase in unification. How does possible-world semantics fare from this point of view? Not very well, I'm afraid. For, as we have seen, the way a theoretical structure acquires unifying power is through its ability to fruitfully interact with other pieces of theoretical structure. But not only are the entities postulated by possible world semantics completely sui generis and unrelated to any theoretical entities we have previously encountered, they are not even in the same 'world' as the latter. It is extremely difficult, therefore, to see how they could relate to the rest of our theoretical structure. [Friedman, 1981, p.13]

The point is not just that we have no reason to take possible worlds seriously, but that a competitor should aim at fulfilling its proper role as a unifier. Goodman's

theory of projection promises to do just that, for he now proceeds to argue that the theory of projection will not only give an account of dispositional terms, but also solve the problem of induction in general. If he is able to point to constraints on the notion of "projectibility", then he may be able to explain why we project some predicates and not others. This will then apply both to understanding the process of prediction (why one prediction and not another) as well as the use of dispositional terms (what is their nature and role in scientific understanding). This would, indeed, conform to the desideratum of connecting with other pieces of theoretical structure that serve to explain other phenomena. The argument for the concept of "projection" lies in its promised unifying power.

Hume's answer to the question how predictions are related to past experience is refreshingly non-cosmic. When an event of one kind frequently follows upon an event of another kind in experience, a habit is formed that leads the mind, when confronted with a new event of the first kind, to pass to the idea of an event of the second kind. The idea of necessary connection arises from the felt impulse of the mind in making this transition. Now if we strip this account of all extraneous features, the central point is that to the question "Why one prediction rather than another?", Hume answers that the elect prediction is one that accords with past regularity, because this regularity has established the habit... And I think his answer is reasonable and relevant, even if it is not entirely satisfactory... [Goodman, 1965, pp.60-1]

Hume's solution is inadequate, Goodman now argues, for

want of that same thing as the theory of dispositional discourse - namely a theory of projectibility. To see this we need to recast the old problem of induction into the "new riddle" of induction.

The real inadequacy of Hume's account lay not in his descriptive approach but in the imprecision of his description. Regularities in experience, according to him, give rise to habits of expectation; and thus it is predictions conforming to past regularities that are normal or valid. But Hume overlooks the fact that some regularities do and some do not establish such habits; that predictions based on some regularities are valid while predictions based on other regularities are not. ... consider our case of emeralds. All those examined before time t are green; and this leads us to expect, and confirms the prediction, that the next one will be green. But also, those examined are grue ["grue" is defined as "being green if examined before time t and being blue if first examined after time t ", where t is some time in the future]; and this does not lead us to expect, and does not confirm the prediction that the next one will be grue. Regularity in greenness confirms the prediction of further cases; regularity in grueness does not. To say that valid predictions are based on past regularities without being able to say which regularities, is quite pointless. ... the problem of prediction from past to future cases is but a narrower version of the problem of projecting from any set of cases to others. We saw that a whole cluster of troublesome problems concerning dispositions and possibility can be reduced to this problem of projection. That is why the new riddle of induction, which is more broadly the problem of distinguishing between projectible and non-projectible hypotheses, is as important as it is exasperating. [Goodman, 1965, pp.82-83]

But when Goodman finally comes to his account of projectibility it is not quite what it promised! He provides little more than a few ad hoc rules concerning

the competition among predicates for the entrenchment of their position in our conceptual frameworks. For instance:

. One principle for eliminating unprojectible projections, then, is that a projection is to be ruled out if it conflicts with the projection of a much better entrenched predicate. [Goodman, 1965, p.96]

Unfortunately, the accompanying account of entrenchment is hardly illuminating. In the grue-green case he says:

Plainly "green", as a veteran of earlier and many more projections than "grue", has a more impressive biography. The predicate "green", we say is much better entrenched than the predicate "grue". [Goodman, 1965, p.94]

At least Goodman is aware of the obvious objection:

New and useful predicates like "conducts electricity" and "is radioactive" are always being introduced and must not be ruled out simply because of their novelty. [Goodman, 1965, p.97]

But he says nothing very insightful about what does rule for or against the "entrenchment" of a predicate, and in a fit of the exasperation, referred to earlier, he says:

The reason why only the right predicates happen so luckily to have become well entrenched is just that well entrenched predicates have thereby become the right ones. [Goodman, 1965, p.98]

At that point, Goodman has not managed to greatly

improve upon the imprecision of Hume's description. Most of what Goodman says about the grue-green problem is right, but it would be nice to place more constraints on entrenchment for other cases at least. In his effort to generalize the problem, Goodman has made it appear less tractable than it really is.

We already had a hint that something was amiss when Goodman said that (1965, p.44) "The problem is to define such [dispositional] predicates [such as "flexible"] solely in terms of manifest predicates [such as "flexes"]." This is reminiscent of the dominant positivist trends in the philosophy of science at the time of Goodman's writing, in particular the program to operationally define every theoretical predicate in terms of observational predicates. All we need to do is construe 'dispositional' as meaning 'theoretical' and 'manifest' as meaning 'observational'. Except that Goodman adds a new twist to the proposed solution. The role of theoretical terms is to "project" the observational terms to a wider class of things. The terms "flexes" and "fails to flex" apply only to the class of material things that have been subjected to a suitable pressure. But the terms "flexible" and "inflexible" apply to the wider class of all material objects.

The extension of "flexes" (the set of things that

satisfy the predicate) is strictly included in the extension of "flexible". But not all cases are like this. For the objective predicate 'red' is not merely an expansion of the manifest predicate 'appears red' to a wider class of things, because some things appear red without being red (e.g. a white object viewed under red light - but anything that flexes must be flexible). Goodman's simple-minded view of the problem is due to the particular examples he chose. The route towards generalizing the problem (and that is important for potential unification) is not in connecting it to tricky problems of induction so much as just looking at a wider range of actual examples more seriously.

The role of dispositional terms in science, I conjecture, is basically the same as for any theoretical vocabulary. All 'objective' predicates are, in some sense, dispositional as we noted earlier. Let us return to the example of size judgements once more and imagine how that might lead to a theory of 'size' as an objective property of things. First we notice that size judgements display a high degree of invariance when referred to the same object viewed under changing circumstances (recognized on different occasions by its distinctive features or viewed continuously over a period of time). The concept of 'size' becomes relatively well entrenched once the perceptual system has achieved some degree of

invariance or stability by adapting to the particular ecological cues in its visual environment. Thus far we have little more than a modern version of the Humean explanation. But now, imagine, that the concept of 'size' begins to play an important role in other observed physical regularities such as Archimedes' principle - the same object will display the same amount of liquid when completely immersed, and when the weight to volume ratio is below a certain value (which depends on the density of the liquid being used) the object will float. The concept of 'size', or at least the concept of 'volume', develops an even higher degree of cross-situational invariance. The concept of 'volume' acquires a wide range of "independent modes of detection", which agree with one another. This is especially so once 'size' is connected with the concept of distance, which is in turn embedded within the deeper space-time theories at the base of our physics.

The manifest predicate 'size' slowly develops into the objective predicates of 'length' and 'volume'. But this evolution is not simply an extension of its use to cover a wider class of things, but very much a refinement of its use - a refinement that is inextricably tied up with theory construction and measurement. And the key feature of the scientific enterprise, it has already been argued, is the search for cross-situational invariance of it

theoretical concepts.

This certainly agrees with Goodman's account in that it is the property of cross-situational invariance that establishes the objectivized essential role in prediction. Consider the example of the mass balance again. The only way that the theory can make a significant prediction is by means of Sneedian type of constraint that says that the mass of the same object is the same for different applications. Without this projection of the theoretically measured value of mass on one occasion to other occasions, no interesting prediction can be made. The Sneedian formalism shows that this is quite a general fact about Newtonian mechanics at least. And we make such projections of 'mass' on the basis of past demonstrations that the 'mass' function is indeed invariant across situations. This the route towards solving Goodman's problem as I see it.

On this view, merely being a veteran of past projections is not sufficient for continued entrenchment. New predicates can displace the veterans by achieving a greater cross-situational invariance in virtue of being embedded in more successful theories. Thus "having a certain surface reflectancy" can replace "green" in scientific discourse in virtue of attaining a greater cross-situational invariance from a successfully verified

physical theory. And 'mass' will replace 'weight' as an intrinsic property of matter in virtue of explaining the invariance of 'weight' to the extent that it is genuine, as well as helping to explain the cases in which the invariance breaks down. Thus far we have succeeded in placing some constraints on what is necessary for the projectibility of a predicate such as 'mass'.

The principle of cross-situational invariance states that for predicate A to replace B it is necessary for A to lead to greater or equal cross-situational invariance than B. Of course this will be a highly complicated function of the theories in which the terms are buried.

But this does not solve the grue-green problem better than Goodman's own account, because the grue predicate is defined in such a way that it inherits exactly the same cross-situational invariance as the predicate "green". This is exactly what gives rise to the puzzle in the first place. On the other hand, it does not violate the principle of cross-situational invariance either, for this only states a necessary condition for continued entrenchment or replacement.

I can't agree with Goodman's anti-realist sentiment that the entrenched predicates simply become the right ones. It is true that we cannot explain the entrenchment of "green" over "grue" by saying that the projection of

"green" will lead to true predictions and the projection of "grue" will lead to false ones. For this would be to explain a present fact as being caused by a future fact — this is unacceptably teleological. But from this trivial observation — that their future reliability does not cause their present entrenchment — it does not follow that their present entrenchment must cause their future reliability. Entrenched predicates do not simply become the right ones in virtue of being entrenched. There is, of course, a strong correlation between "being entrenched" and "being reliable" but the explanation of that correlation will become a non-trivial part of a future naturalized epistemology. As Quine pointed out, that explanation will rely on the very theories whose success we seek to explain. But this circularity does make the explanation trivial or vacuous.⁹

1.5.3 'Causality' as a Theoretical Concept

It would have seemed trivial to have stated earlier that 'causality' is just another theoretical concept. But now this assertion takes on added meaning. The role of all theoretical concepts that are more than purely representational is to establish cross-situational

9. Stabler, 1982, pp.166-7, argues against Putnam's charge to this effect — that evolutionary epistemologies are necessarily vacuous.

invariances. So this, too, should be the role of 'causality' in scientific discourse if it is to have any significant role at all. This throws a completely novel light on traditional philosophical discussions of the concept.

Traditionally, the "trouble with causation" has been seen as similar to that plaguing dispositional terms in general, because there are similar "counterfactual translations". "Event a causes event b" seems to translate to, roughly, "a necessitates b" where this is taken to modally imply not only inevitable succession of events of these types but also that "if a had occurred then b would have occurred" even in a case in which not-a is true.¹⁰ Quine is typical in describing the problem:

The trouble with causation is, as Hume pointed out, that there is no evident way of distinguishing it from mere invariable succession. And why is this troublesome? Because then, if we take any two classes of events such that each event in the one class is followed by an event in the other, we have to say that the events in the one class cause those in the other. Thereupon any arbitrary event a can be said to have caused any succeeding event b; for, we can just take the two classes as the unit classes of a and b... There is the same quandary over dispositions. If there is no distinguishing between a thing's disposition to act in a certain way in certain circumstances and the mere fact of its so acting in those circumstances, then whatever the thing may do can be laid down to a disposition, by defining

10. The development of these analyses of causation is described in chapter 2.

the circumstances narrowly enough. [Quine, 1974, p.5]

The last question that Quine raises about dispositions in general - how do we avoid treating every dispositional assertion as being "grounded" in reality - has already received an answer. We only take such terms seriously when they successfully achieve a degree of cross-situational invariance. Until that time they subsist in everyday language, like parasites, hoping to latch onto a host theory. But the question is now being asked of 'causality'. How successful is this parasite?

The paradigm case of causation that Hume discussed was the example of two billiard balls. Ball A hits ball B and causes it to move. Hume asked what the necessity of this sequence of events consists in and what evidence we can have for that necessity. If my conjecture is right, then the concept of 'cause' will only play a role in explaining lower level phenomenological regularities if it helps establish higher order cross-situational invariances that transcend those in the explanandum (that which is being explained).

Indeed the billiard ball example supports this picture. The observed regularity - on numerous occasions when ball A hits ball B it moves - does not hold universally. If we interfere in certain ways, such as placing a finger on B or by shooting a third ball towards

B from the opposite direction, then the "regular succession" breaks down. So the useful work done by a concept of 'cause' is to transcend such boundaries and claim even these recalcitrant cases as within its explanatory domain. In this way it will explain not only the observed regularities but also its failure. In fact, Quine's own view of causation supports this conjecture, because he sees the theoretical concept of energy as the central idea.

The imparting of energy seems to be the central idea [behind the causal idiom]. The transfer of momentum from one billiard ball to another is persistently cited as a paradigm case of causality. Thus we might seek the simpleminded or root notion of causality in terms of the flow of energy. Cause and effect are events such that all the energy in the effect flowed from the cause. [Quine, 1974, p.5]

What other reason can there be for analysing 'cause' in terms of the theoretical concept of 'energy'? "a causes b" means for Quine "energy flows from a to b". This exactly safeguards against the examples mentioned earlier. In such cases of interference the kinetic energy in A at impact is not transferred to B but is re-absorbed by A, which proceeds to move in the opposite direction with the same energy as before. The concept of 'energy' helps explain the phenomena in a wider range of cases.

We are now in a position to reconsider van Fraassen's use of Reichenbach's principle of common cause as the best

explication of the realist's methodological stance. First of all, it is not clear that Reichenbach's principle captures the notion of 'cause' in a useful way at all. (In chapter 2, I will argue that it does not. In chapter 3 I will develop an account of causality that stands up to such scrutiny.) Secondly, the realist account of explanation and prediction finds a deeper formulation in the principle of cross-situational invariance, as has been sketched earlier.

The search for "hidden common causes" is just one strategy that may lead to greater cross-situational invariance of theoretical concepts, which can then be used for explanation or be projected into other applications in order to make predictions.

It is possible to see that the principle captures the basic idea behind Reichenbach's Principle of Common Cause, at least in the cases in which it applies. As an example, suppose that we see that two television sets display the same patterns of light ² over different occasions in time. How do we explain this correlation? It might appear that we could just say that one event causes the other. We don't say that exactly because the correlation is not always invariant. If we interfere with one of the sets, say by changing the channel, then we see that the correlation no longer obtains. We need a hypothesis to

explain the correlation under the conditions that it does obtain and to explain why it does not obtain under other conditions. The actual explanation is in terms of common cause - there is one signal sent from a certain transmitter that is received by both television sets when their tuners are set to resonate at the same frequencies. When they are not so set, a common signal is not received. The strength of this explanatory story lies in the fact that something common to both types of situations is found, namely the existence of electromagnetic signals at both locations originating from a common source. Again the underlying story is theory bound. The explanation that would explain the correlation in terms of direct causal influence between the television sets is deemed mystical in this case exactly because it lacks any cross-situational coherence.

The motivation behind the Principle of Common Cause, correctly understood, accords exactly with the realist's desire not to leave 'cosmic coincidences' unexplained. But the Principle of Cross-Situational Invariance is far more general. It explains why we should not posit a causal relation too liberally, and hints at what counts as a good explanation even when a common cause relation cannot be assigned, as in the quantum mechanical case. (That will be discussed further in chapters 4 and 5.)

The purpose of this chapter has been to outline the broader research program in which the following more detailed chapters are embedded.

Chapter 2

IN DEFENSE OF THE INTERVENTION THEORY OF CAUSALITY

2.1 Introduction

One aim of this chapter is to highlight some of the problems with the "traditional" probabilistic theories of causality (section 2) - e.g. Suppes (1970), Salmon (1970), Cartwright (1979), Skyrms (1980), Eells & Sober (1983), Bunzl (1984), as well as different approaches such as Lewis (1979). All of the probabilistic theories rely on some interesting properties of partial regression coefficients. Unfortunately there are deep difficulties with this approach. In particular, there is a puzzle concerning deterministic examples and secondly we have the age-old problem of doing justice to the temporal asymmetry of cause without introducing ad hoc assumptions. These objections affect the different analyses differently so they will be treated individually as required in section 2. Subsection 2.2.3 discusses Lewis' counterfactual analyses of causation, while the subsequent subsection criticizes Cartwright's (1979) paper as a strong

representative of the probabilistic¹ approach. Bunzl's (1984) paper is important for my purposes as it is the only philosophical paper (as far as I'm aware), except Rescher & Simon (1966), that deals at length with the ideas behind Wright's theory of path analysis. This paper is scrutinized in sections 2.3 and 2.5.

Both of these problems will be solved in the process¹ of securing new foundations for Wright's path analysis in terms of the "intervention" or "manipulability" theory of causation. The view of causality thereby developed handles the results proved in a recent paper by Eells & Sober (1983) on the question of the transitivity of causal influence in a far more perspicuous manner, as just one example (chapter 3, section 3). The metaphysical status of causality resulting from this view is handled by the main result of section 3, which proves that the coefficients that measure causal influence on the "intervention" view can be equated with the constant parameters characterizing the system, thus effecting a satisfying ontological reduction. The epistemological aspects are handled by the inverse theorem in section 6, where it is shown that assumptions concerning the robustness of statistical data

1. See Wright, 1968, for a concise exposition and references - Wright is best known as the co-founder of modern population genetics - along with Fisher and Haldane.

under various conditions of intervention form the correct grounds for causal inference on the intervention view.

More generally I am concerned with the import of scientific realism on the interconnections among laws, counterfactual conditionals, probability, and causality. In this regard, it is in the spirit of the realist view of laws of nature developed recently by Dretske (1977), Armstrong (1978), and Tooley (1978). The noticeable difference with my approach is that I start from the mathematical representation of nature. The advantages I claim here are roughly: (1) there is no presupposition that scientific laws are reconstructable in terms of artificial logical languages, (2) there is no problem about explaining the logical relations between laws (qua relations between universals) and true universal generalizations [see Armstrong (1982) and Lewis (1983)] whereas the relationship between basic and derived equations is clear, and (3) that the distinction between derived and basic equations is seen as interesting and useful for describing scientific practice, e.g. reductionistic methodology as a "search for basic equations", whereas the corresponding distinction in the Dretske-Tooley-Armstrong account has no obvious relevance to real science.

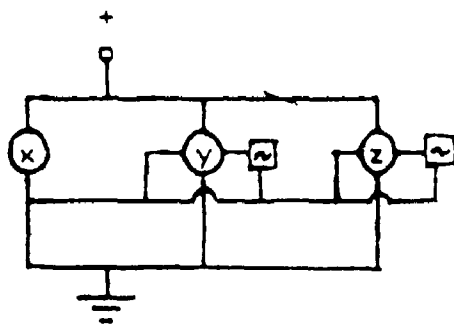
2.2 Two Problems with Probabilistic Causality.

2.2.1 The Problems

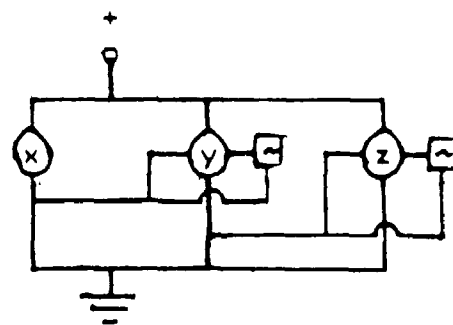
Suppose we have 3 event-types represented by random variables X , Y , and Z , where each of X , Y , and Z is a dichotomous random variable taking on values 1 or 0. Tichý (1978) presents an interesting framework in which to construct examples. Imagine that we have 3 relay switches labelled x , y , and z , each of which is either on or off on any occasion. $X=1$ stands for the event of x being on and $X=0$ means that x is off. Similarly for Y and Z . They operate on a discrete time scale, t_0, t_1, \dots, t_n . A switch is on at time t_k when it has received an electrical pulse from the left at time t_{k-1} , and it is off at time t_k if it received a pulse from the right at t_{k-1} . I will take the random variable X to refer to the state of the switch x at time t_0 , Y to switch y at t_1 and Z to z at t_2 . Other choices are possible, but the important point to notice is that none of what follows makes much sense unless each random variable is taken to be indexed to some particular time.

We can easily imagine two sorts of circuit mechanisms that fit cases (a) and (b). For example, consider the

circuits diagrammed below, where the box with the \sim (standing for negation) produces a positive output with a negative input and vice versa.

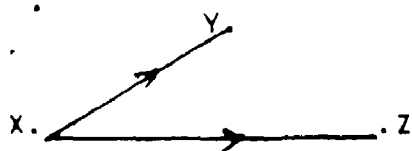


(a) Common causation.



(b) Serial causation.

The causal relations that hold among the events X , Y , and Z in these two cases are; (a) common causation, or (b) serial causation, as illustrated diagrammatically below;



(a) Common causation.



(b) Serial causation.

A different example that fits (a) would be two television sets tuned to the same channel. $Y=1$ might indicate that the brightness of the picture for the first TV set is above a certain threshold, whereas $Y=0$ otherwise. Similarly, $Z=1$ and $Z=0$ indicate the state of TV_2 . Each state is determined by the type of signal emitted at the transmitting station, of type indicated by $X=1$ or $X=0$. On the other hand the situation depicted in

(b), above, would be exemplified when TV_2 , instead of being tuned to the same channel as TV_1 , displays the same picture as TV_1 in virtue of the fact that scintillations are electrically encoded and relayed via cable to TV_2 . Under hypothesis (a), the events $Y=1$ and $Z=1$ are each caused by the event $X=1$ (the common cause). Under hypothesis (b), $Z=1$ is caused by $Y=1$, which was caused by the event $X=1$. The task of any satisfactory probabilistic theory of causality is to distinguish between these two cases on the basis of probabilistic regularities.

The 'traditional' theories, as I will call them, attempt to do this roughly in accordance with the following definitions;

I: The event $Y=1$ has a direct (positive) causal influence towards the occurrence of the event $Z=1$ if, and only if, $R_X(Z/Y) > 0$, where

$$\begin{aligned} R_X(Z/Y) &= P(Z=1/X=1.Y=1) - P(Z=1/X=1.Y=0) \\ &= P(Z=1/X=0.Y=1) - P(Z=1/X=0.Y=0) \end{aligned}$$

II: The event $X=1$ has a direct (positive) causal influence on $Z=1$ if, and only if, $R_Y(Z/X) > 0$, where

$$\begin{aligned} R_Y(Z/X) &= P(Z=1/X=1.Y=1) - P(Z=1/X=0.Y=1) \\ &= P(Z=1/X=1.Y=0) - P(Z=1/X=0.Y=0) \end{aligned}$$

The quantities labelled $R_X(Z/Y)$ and $R_Y(Z/X)$ are called

partial regression coefficients because they are regression coefficients with the subscripted variable held fixed. The assumption that it doesn't matter which value the subscripted variable, or variables, is held fixed is known in the literature as the "no-interaction" assumption and will be shown later to be equivalent to a linearity assumption. The example Cartwright gives of "interaction" is the case of "ingesting an acid poison [$Y=1$] may cause death [$Z=1$]; so too may the ingestion of an alkali poison [$X=1$]. But ingesting both will have no effect on survival." [Cartwright 1979, pp.427-8]

The intuitive motivation behind these definitions can be presented as follows. In definition I we are considering the difference in the probability of $Z=1$ between two situations in which the value of Y is different but the value of X is the same. So, intuitively, any difference in Z can only be attributed to the difference in Y , because X is the same in both cases. Similarly $R_v(Z/X)$ measures the difference in the probability of $Z=1$ that can only be attributed to a difference in X , since Y is held fixed. Conversely, suppose that the common cause situation is true, that events at y have no causal influence on what happens at z , but there is nevertheless a strong correlation between what happens at y and what happens at z : $Z=1$ occurs almost always in association with $Y=1$ and $Z=0$ almost

always with $Y=0$. The association arises because $X=1$ tends to bring about $Y=1$ as well as $Z=1$, and $X=0$ tends to bring about $Y=0$ as well as $Z=0$. As a consequence of these regularities we have that $P(Z=1/X=1.Y=1)$ is close to 1 and $P(Z=1/X=0.Y=0)$ is close to 0, but this alone tells us very little about the size of the differences as defined by $R_x(Z/Y)$ and $R_y(Z/X)$. The crucial information is obtained when $Y=0$, say, despite the fact that $X=1$. In this (rare) case we would expect that $Z=1$ is still most likely, i.e. $P(Z=1/X=1.Y=0)$ be close to 1, in the common cause case because it is the state at x that really matters, so that $R_x(Z/Y)$ is close to 0 and $R_y(Z/X)$ is close to 1. But we expect $P(Z=1/X=1.Y=0)$ to be low (close to 0) in case that Y strongly influences Z because then it is the value of Y that matters in determining the likelihood of Z being 1. In that case we would have $R_x(Z/Y)$ close to 1 and $R_y(Z/X)$ close to zero. So the values of the partial regression coefficients differ noticeably for the two cases, as was hoped. The point to notice, however, is that it is the information about what happens in the rare cases when the regularity between X and Y is "broken" that tells us whether it is X or Y that is causally determining the value of Z .

So the first problem that arises for this 'traditional' analysis is when these rare cases are so rare that they never occur! In that case the partial coefficients

$R_x(Z/Y)$ and $R_y(Z/X)$ are ill defined because one of the conditional probabilities appearing there has no value that can be determined from Kolmogorov's definition of conditional probability.² As a concrete example, suppose that $P(X=1, Y=1, Z=1) = 1/2 = P(X=0, Y=0, Z=0)$ and all other states of affairs have probability 0. These regularities are consistent with either of the causal hypotheses (a) or (b) above. The values of all three variables, X, Y, and Z, would co-vary in exactly the same way in both cases! The problem is, in other words, that the phenomenological regularities, probabilistically described, underdetermine³ the causal facts in any deterministic example.

Besides this, there is another objection to traditional "partial regression" approach to probabilistic causality that is, perhaps, even more telling. It is the old problem of accounting for the temporal asymmetry of causal action in a non ad hoc manner - a problem that has plagued causal logics from the beginning. In the common cause model; X influences Z, but Z does not causally influence X. Unfortunately the expression obtained from definition

2. In the classical Kolmogorovian theory of probability conditional probabilities $P(A/B)$ are defined as $P(A.B)/P(B)$ and this is 0/0 in case $P(B) = 0$.

3. A deterministic example is, roughly, one in which all conditional probabilities are 0 or 1. An indeterministic case is one in which some conditional probabilities are strictly between 0 and 1.

II by replacing X with Z and Z with X , that is $R_V(X/Z)$, is greater than zero if and only if $R_V(Z/X)$ is greater than zero. So that Z causes X exactly when X causes Z , just as a theorem of probability theory? To get around this unpalatable result one must add the assumption that Z cannot act on X because the events $X=x$ occur after $Z=z$ in time. This is exactly what Suppes (1970, p.25) does, and what Salmon has been criticized for not doing (Shrader, 1977). Adding the condition that the present can only be caused by the past makes it analytically true that there is no backwards causation, but surely this is an empirical matter! The weaker assumption that there is some fixed causal ordering is at once too weak and too strong. Too weak because there are no rules given for saying what the ordering is in any given case, and too strong in that it rules out even the possibility of two-way causation.

Both of these ad hoc assumptions - the meaningfulness of all conditional probabilities and the temporal ordering imposed upon causation - are eliminated as superfluous in the intervention account of causality to be developed here. This is an eminent virtue for any analysis of causation.

2.2.2 The Solution in Brief

What does distinguish between these two causal stories,

(a) and (b) above, is what would happen were we to impose the condition $Y=0$ on the system from outside, without changing the fact of $X=1$, say. If we find that $P(Z=1/[Y=0].[X=1]) = 1$, where the square brackets indicate that the condition is imposed on the system from outside, then it is hypothesis (a) that holds and not (b). So taking the notion of imposition as primitive for the moment, the improved analysis that I propose is:

Definition 1:

(i) The direct causal influence of $Y=1$ on the occurrence of $Z=1$ is measured by $r(Z/[Y])$, where

$$\begin{aligned} r(Z/[Y]) &= P(Z=1/[X=0][Y=1]) - P(Z=1/[X=0][Y=0]) \\ &= P(Z=1/[X=1][Y=1]) - P(Z=1/[X=1][Y=0]) \end{aligned}$$

(ii) The total causal influence of $Y=1$ on the occurrence of $Z=1$ is given by $R(Z/[Y])$, where

$$R(Z/[Y]) = P(Z=1/[Y=1]) - P(Z=1/[Y=0])$$

Similar definitions hold for causal influences between any two events. The measures of causal influence used in the above definitions will be referred to as causal coefficients. The difference between direct and total causal influence is that to obtain the direct influence we must hold all possible intermediary causes fixed while we "compare" what happens to Z when we vary Y . When we don't

hold anything fixed, as in definition (ii), then we are allowing Y to influence Z indirectly by firstly influencing X , which in turn influences Z , so in that case we are measuring total causal influence.

The definitions proposed above solve the problem for deterministic examples because all probabilities are well defined. We don't have to wait for rare cases to occur naturally (you have to wait for ever in deterministic cases), but rather we extract the information we need by manipulating the system. And there is an additional bonus! The temporal asymmetry problem is solved as well. We don't want it to follow logically that Z causally influences X from the fact that X causally influences Z . And this indeed comes out right, because $r(X/[Z]) > 0$ does not follow from $r(Z/[X]) > 0$, as one would hope.

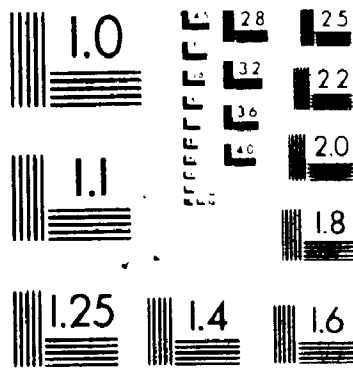
2.2.3 Counterfactual Analyses of Causation

The traditional probabilistic analyses of causation reduce to traditional counterfactual analyses of causation in the extreme case where $R_V(Z/X) = +1$. For each conditional probability lies between 0 and 1 so $R_V(Z/X) = +1$ if, and only if

$$(i) P(Z=1/X=1, Y=y) = 1, \text{ and } (ii) P(Z=1/X=0, Y=y) = 0$$

where y can be 0 or 1, according to the no-interaction

2



assumption. The probabilistic assertions can easily be seen to correspond to the counterfactual conditionals

$$(i) X=1 \rightarrow Z=1, \text{ and } (ii) X=0 \rightarrow Z=0.$$

where " \rightarrow " stands for "if it were the case that ... then it would be the case that ...". This is basically the earlier analysis discussed in Lewis (1975). In accordance with the usual treatment of counterfactuals [see the Ifs volume, Harper (1981)], a conditional is true just in case the consequent is true in the nearest possible world(s) in which the antecedent is true. The content of the analysis lies in the specification of what counts as "the nearest world".

The traditional counterfactual analyses have the immediate disadvantage of not being applicable to
4
indeterministic cases. But the close correspondence with the traditional probabilistic theories in the extreme deterministic case shows that they are also subject to the same problems as presented earlier.

How does the first problem arise? The no-interaction assumption says that it does not matter whether $Y = 1$ or 0 in the possible world(s) in which the counterfactuals are

4. But this defect might be eliminated by generalizing to conditionals with chance consequences as in Skyrms (1980).

evaluated, as long as we are consistent. It would be a serious mistake to evaluate (i) in a world in which $Y=1$ and (ii) in one in which $Y=0$. But in the deterministic case, there are no possible worlds in which $X=0$ and $Y=1$ or in which $X=1$ and $Y=0$, so the two conditionals cannot be evaluated in any other way. This is exactly the confusion that is invited by the counterfactual analyses by not making the conditions under which the conditional is to be evaluated explicit. The probabilistic approach has the advantage of not encouraging this type of vagueness.

The "time asymmetry" problem is simply the problem of how to disallow the rule of contraposition for counterfactual conditionals. Lewis (1979) addresses this problem by proposing that the counterfactual conditionals be understood as non-backtracking. This means, according to Lewis, that we are to imagine a possible world in which the antecedent is brought about "by means of a small miracle". Because this means that we don't have to "backtrack" to provide conditions under which the antecedent would or could arise naturally, the "nearest possible world" is uncontroversially one in which no events, except the counterfactual antecedent, is different from the actual world. Lewis addresses the "time asymmetry problem" but inadvertently solves the other problem I have raised as well - but at the expense of introducing "miracles" into the analysis. I don't believe

in counterfactual miracles no matter how small. How could we ever test the truth of what would happen were there miracles if there never are any?

For me the antecedents are imposed from outside the system, so that they are only "miracles" in the innocuous sense that relative to the system they are uncaused, i.e. the events that do cause the intervention have no causal connection to the system other than via the intervention.⁵ This system relativization of the backtracking/non-backtracking distinction is, I think, enlightening, and it is the only sensible way to handle the vagueness inherent in the use of counterfactual conditionals. Of course, in order to carry out my program, I need to eliminate the vagueness inherent in the word "system". To this end I will always characterize any particular system mathematically in terms of some set of equations, as is customary. The next section deals with what inferences concerning probability, cause, and

5. The idea that outside intervention is crucial to the analysis of causality appears in Rescher & Simon (1966). There are hints that this had been recognized by Wright, whose work on the subject goes back to early in this century. Wright talks about "imposed constancies" but it is not clear that he means this in a physical sense, since he fails to make the distinction between the causal coefficient $r(Z/[X])$ and the partial regression coefficients $R_v(Z/X)$ crucial to my work. Indeed the usual treatments of path analysis and causal modelling start with an assumed causal ordering among variables and use a partial regression analysis to determine the weights associated with each "path".

intervention can be drawn from such premises.

2.2.4 Cartwright's Analysis

In a classic publication, Cartwright (1979) took several earlier analyses of causality to task and provided her own. Let me just state her definition and then explain it in terms of the simple three-variable examples above.

CC: $C \rightarrow E$ [C causes E] iff $P(E/C, K_j) > P(E/K_j)$ for all state descriptions K_j over the set $\{C_i\}$, where $\{C_i\}$ satisfies

- (i) $C_i \in \{C_i\}$ implies $C_i \rightarrow \pm E$
- (ii) $C \in \{C_i\}$
- (iii) For all D ($D \rightarrow \pm E$ implies $D=C$ or $D \in \{C_i\}$)
- (iv) $C_i \in \{C_i\}$ implies $\sim(C \rightarrow C_i)$

As Cartwright says, this is only a constraint on the meaning of the causal relation as the schema $C \rightarrow E$ appears on both sides of the equivalence. Intuitively, the set $\{C_i\}$ contains all variables that causally influence C, and these are held fixed in the 'state descriptions' K_j . Note that $P(E/C, K_j) > P(E/K_j)$ is equivalent to

$$P(E/C, K_j) - P(E/\neg C, K_j) > 0$$

if $P(C/K_j) > 0$, so we see that the analysis makes use of

partial regression coefficients in a restricted sense. Condition (iv) ensures that causes intermediary between C and E are not "held fixed" so that this analysis is of indirect causation.

The strength of the analysis is that it avoids the difficulty of building in a causal asymmetry, but at the expense of weakening the constraint on the concept of cause. The difference between a definition of 'cause' and an analysis of 'cause' is worth noting here. The concept of 'cause' is fairly primitive and the hope of reducing it to more fundamental notions is not likely. The aim of analyses of 'cause' are therefore more modest. Here we only aim to place constraints that tie the notion down, either to itself or other concepts such as probability and nomic regularities. The success of any such analysis should be judged as for any theory - in terms of explanatory unification, lack of adhocness, parsimony, and heuristic potential. By these criteria Cartwright's analysis fares better than Suppes (1970). For this reason I have chosen to neglect to discuss Suppes' theory because it is adequately ~~scrutinized~~ scrutinized in Cartwright (1979) and by Otte (1981).

Although Cartwright's theory has many virtues, it does not cope with the problem of deterministic examples. In terms of our notation, Cartwright's CC comes to: $Y \rightarrow Z$ iff

$P(Z=1/Y=1, X=1) > P(Z=1/X=1)$ and $P(Z=1/Y=1, X=0) > P(Z=1/X=0)$, because $\{C_1\} = \{X=1, X=0\}$. The second conjunct contains an undefined probability because $P(Y=1, X=0)=0$. Cartwright addresses the problem of 0-1 probabilities on page 428 by changing CC. In terms of the present framework the new "definition" is:

$C \rightarrow E$ iff (1) $P(Z=1/Y=1, X=1) > P(Z=1/X=1)$ or $[P(Z=1/Y=1, X=1)=1=P(Z=1/X=1)$ and $(P(Z=1/X=1) < 1$ or $P(Z=1/X=0) < 1)]$ and (2) $P(Z=1/Y=1, X=0) > P(Z=1/X=0)$ or $[P(Z=1/Y=1, X=0)=1=P(Z=1/X=0)$ and $(P(Z=1/X=1) < 1$ or $P(Z=1/X=0) < 1)]$

Notice that the previous virtue of lacking adhocness has now been lost. But even worse, it fails to solve the problem. For consider the deterministic versions of cases (a) and (b). In both cases it turns out that Y influences Z, but this is the wrong answer for the common cause case (a). Of course, this is hardly suprising because both cases, (a) and (b), are probabilistically identical.

This same defect has been noted in Otte's (1981) critique of Suppes (1970).

The basic problem facing a definition of spurious causation which depends solely on probability relations among events seems to be that there is no way to distinguish between a causal chain and a causal fork when the probabilities are the same. Suppose that we have the chain of necessary and sufficient causes $C \rightarrow A$ and a necessary and sufficient common cause $C \rightarrow A$. All the probability

relations in the two cases will be identical, yet in one case B is a genuine cause of A and in another case B is a spurious cause of A. Thus there appears to be no way to distinguish genuine from spurious causes using only probability relations among events, which is a very serious problem for a probabilistic theory of causality. [Otte, 1981, p.180.]

It is significant to observe, therefore, that the very same objection also applies to Cartwright (1979), even though it is not purely probabilistic in the sense that a restricted set of causally relevant background factors is singled out by non-probabilistic means. So even with this weakening, Cartwright has not solved one of the major objections to Suppes' analysis.

2.3 Objections to the Intervention Account of Causality

In a recent publication, Bunzl (1984) has discussed the problem of "causal asymmetry" or "causal priority" as he likes to call it. Like myself, Bunzl thinks that equating causal priority with temporal order is unsatisfactory. The problem he then addresses is stated as "...if time plays no part, can we distinguish between the direction of causal relations between two events that are assumed to be causally related?"

One tempting way to do so seem to be by way of an Act or Recipe approach to causation:

that is, by way of an account in which we manipulate some things and observe for "results." Such accounts were originally offered as a way of defining the causal relation. But even if applied to a much more modest (epistemological) role, such an approach begs an important question; namely how are we to distinguish between cases where we are manipulating something and cases where we simply think we are doing so but are in fact being affected by that which we think we are affecting (See Bunzl (1980)). The approach I want is quite different. [Bunzl, 1984, p.32]

I will discuss the approach he takes in a moment, but first I want to address Bunzl's important worry about whether his thirst for coffee will lead him to turn the kettle on or whether the fact that the kettle will soon be on has caused the craving for coffee he now has! To be fair I looked up the earlier note cited.

There, he gives an argument to the conclusion that the notions of agency or action (or simply "doing") can be of no use in "the key problem of fixing the asymmetry of causal relations". The first premise of this argument is stated as follows:

For an appeal to our doing A to be useful in identifying the direction of causal asymmetry between A and B we surely better be able to differentiate between cases when we actually are doing A and cases when we think we are doing A but are in fact undergoing A: that is cases where we think we are acting when in fact we are being acted upon. For unless we differentiate between cases of doing and undergoing, appeals to "doing" will hardly offer a reliable indicator of causal relations. [Bunzl, 1980, p.629]

My reply to this is quite simple: We can differentiate

between doing and undergoing because in the case of undergoing none of the normal internal regularities of the system would be different. Take the relay switch examples, either (a) or (b) it doesn't matter which. Normally y is on at t_1 when x is on at t_0 and y is off at t_1 whenever x is off at t_0 . Suppose we use the manual switch to manipulate the state of y at t_1 . Bunzl asks how we "could tell" if it was the system that was actually manipulating us. Well, the difference is that if the system manipulates us then the normal regularities described above will continue unimpeded. But if we are manipulating the system, then there will be statistical changes there - in particular we will sometime see x off at t_0 followed by y being on at t_1 , and vice versa.

Just in case his challenge might be answered, Bunzl continues.

Whatever these differentia are, they will only be useful indicators of genuine cases of doing if we can differentiate between cases when they actually obtain rather than merely seeming to obtain. [Bunzl, 1980, p.630]

Can we differentiate between cases when x being off at t_0 co-occurs with y being on at t_1 ? If that's the question then the answer is 'yes' again - all we need is to place a lightbulb in the circuit so we can tell when that switch is closed. Of course Bunzl is free to raise further skeptical worries about these points, but he owes us an

account, I think, of why these should be any more damaging for the manipulability account than, say, his own theory.

Sosa (1975) makes a similar but slightly different point when commenting on von Wright's article:

Von Wright then defines causation by reference to the interference of agents: p is a cause relative to q, and q is an effect relative to p, if and only if, by doing p we could bring about q or by suppressing p we could remove q or prevent it from happening.

It seems clear that the definiens here is not to be interpreted so as to mean simply that it is logically (physically) possible that we be bring about q by doing p, etc. An equally idiomatic reading seems more plausible: 'by doing p we could bring about q' may be read as 'if we were to do p, we would thus bring about q'. This does not require that we should be able to do p here and now, so the reign of causation can range far beyond the actual reach of agents.

Some difficulties appear discernible. For instance, is not the proper conceptual order reversed? Is not causation essential in understanding what it is for one to bring about q by doing p? ...as it stands the definition seems open to the objection that one brings it about that one obeys the traffic laws by stopping at the red light, but one's obedience of the traffic laws is not caused by one's stopping at the red light. [Sosa, 1975, pp.7-8]

I take it that the worry is that, when asked what "doing" or "intervention" is, von Wright must use the concept of cause in his reply. I find Sosa's worry rather confusing. First, I agree that 'cause' is the more primitive notion and that we will never understand action

without some appeal to 'cause'. But Sosa thinks that von Wright claims to have reduced 'cause' to 'action' and, in light of the foregoing point, von Wright must be mistaken. No, Sosa is mistaken if he thinks von Wright's analysis makes any claim about reducing 'cause' to 'action'. But what von Wright's definition does claim to do is reduce 'cause' to 'what would happen if an action were performed' or to 'counterfactual action' if you like. And that is something very much weaker and more metaphysically acceptable than 'reducing 'cause' to 'action''. Sosa, I believe, has fallen into the trap of conflating the epistemological and ontological aspects of counterfactual conditionals. That is, he has confused the way in which we may verify these conditionals with that element of reality in virtue of which they are true. We need 'action' for the former, but 'action' has nothing to do with the latter. Von Wright himself addressed exactly this point:

The second misunderstanding is a confusion between the epistemic and the ontic aspect. By no means have I wanted to maintain that the operation of a cause always results from action. Causation, needless to say, operates throughout nature independently of agency, also in regions of the world forever inaccessible to human interference. But the test-procedures characteristic of causal laws, including those whose operation is far removed from us in space and time, belong to the scientists' laboratories - and they belong there essentially, because of their conceptual connection with the mode of action we call experiment. [von Wright, 1975, p.107]

Rather than belabouring these points, I will try to make sense of what Bunzl sees himself as achieving. Back in his 1984 paper, then Bunzl expresses his aim as being to "explicate what causal priority is".

This ... ontological concern ought to be distinguished from epistemological considerations about causal priority; namely how we can tell what the direction of causal influence is between causal relata. Those concerned about the possibility of backward or instantaneous causation have had to worry about both ontological and epistemological considerations... Of course it would be nice if the ontological and epistemological could go hand-in-hand, but my belief is that they don't. My reason for saying that is that I think that an adequate ontological account of causal priority involves modal considerations (see Bunzl, 1984b). And in the argot of modal analysis, possible worlds (even if real) are not worlds from which we can glean much epistemological value. But if we define causal relations by way of possible worlds, are there any circumstances under which we can expect to tell what causal relations actually obtain? [Bunzl, 1984, pp.31-2]

Bunzl sets himself the task of answering the last question. That is, he sets himself the task of giving a list of assumptions that, together with empirical data, are sufficient to fix the direction of causation. Under the (limited) circumstances defined by those assumptions, we can "tell that causal relations actually obtain." The fact that these circumstances are limited is why Bunzl wishes to divorce the epistemological questions completely from the ontological ones.

Because Bunzl is working within the general framework

of path analysis, it is best to introduce that in the next section and postpone the further discussion of his view to section 2.5.

2.4 Path Analysis

Path analysis is founded upon the observation that the constant coefficients appearing in a network of linear equations can be related to various measures of correlation under probability distributions consistent with those equations. We will consider the system of equations (2.1), (2.2), and (2.3) as follows.

$$X = r_{x/y}.Y + r_{x/z}.Z + E_x \quad \dots (2.1)$$

$$Y = r_{y/x}.X + r_{y/z}.Z + E_y \quad \dots (2.2)$$

$$Z = r_{z/x}.X + r_{z/y}.Y + E_z \quad \dots (2.3)$$

where the r 's are constants, and E_x , E_y , E_z are "error" or "noise" terms representing a random input into the system. It is normally assumed that these terms are stochastically independent of the other input terms with which they appear. Thus E_x is independent of Y and Z , E_y is independent of X and Z , and E_z is independent of X and Y .

For any joint distribution defined for these three random variables, X , Y , and Z , the expected value, denoted

$$R(Z/X) = r_{z/x} + r_{z/y}.R(Y/X) \quad \dots (3.3)$$

$$R(Z/Y) = r_{z/y} + r_{z/x}.R(X/Y) \quad \dots (3.4)$$

$$1 = r_{z/x}.R(X/Z) + r_{z/y}.R(Y/Z) \quad \dots (3.5)$$

The question now is by what rules do we read off the values of $R(Z/[X])$, $r(Z/[X])$, etc. The idea here is that the counterfactual situation in which the value of Y , say, is imposed is one in which the equation for Y is inoperative. The usual causal ancestry of Y is no longer efficacious. Instead, the value of Y is determined by some different chain of events unconnected with the system (as defined) - that is what is meant by imposing the value of Y from outside the system. This will be summarized by two rules, but first we need to introduce some notation. Denote the original system of equations governing X , Y , and Z by $S(X,Y,Z)$, and denote the set of equations governing the system modified by the external imposition of certain variables by enclosing those variables in square brackets, e.g. $S([X],Y,[Z])$ denotes the set of equations operating when the values of X and Z are imposed from outside. The question is what is the set of equations governing such modified systems?

Rule(1): The modified system of equations, $S(X,[Y],Z)$ say, is obtained from the original set $S(X,Y,Z)$ by deleting the equation governing any variable in square

constants. If we take expected values of both sides of (2.3) first and then multiply by $\langle X \rangle$, we have

$$\langle Z \rangle \cdot \langle X \rangle = r_{Z/X} \cdot \langle X \rangle \cdot \langle X \rangle + r_{Z/Y} \cdot \langle Y \rangle \cdot \langle X \rangle + \langle E_Z \rangle \langle X \rangle \quad [B]$$

Now using the definitions of covariance and variance,

$$\text{Cov}(X, Y) \stackrel{\text{def}}{=} \langle X \cdot Y \rangle - \langle X \rangle \cdot \langle Y \rangle, \quad \text{Var}(X) \stackrel{\text{def}}{=} \text{Cov}(X, X) \quad \dots (2.5)$$

we have, by subtracting [B] from [A],

$$\text{Cov}(Z, X) = r_{Z/X} \cdot \text{Var}(X) + r_{Z/Y} \cdot \text{Cov}(Y, X)$$

where we have used the assumption that $\text{Cov}(E_Z, X) = 0$. In fact we can obtain 3 such equations from (2.3), namely;

$$\text{Cov}(Z, X) = r_{Z/X} \cdot \text{Cov}(X, X) + r_{Z/Y} \cdot \text{Cov}(Y, X) \quad \dots (2.6)$$

$$\text{Cov}(Z, Y) = r_{Z/X} \cdot \text{Cov}(X, Y) + r_{Z/Y} \cdot \text{Cov}(Y, Y) \quad \dots (2.7)$$

$$\begin{aligned} \text{Cov}(Z, Z) &= r_{Z/X} \cdot \text{Cov}(X, Z) + r_{Z/Y} \cdot \text{Cov}(Y, Z) + \text{Cov}(E_Z, Z) \\ &\dots (2.8) \end{aligned}$$

In the deterministic case, in which E_Z is constant (and therefore $\text{Cov}(E_Z, Z) = 0$), these equations can be seen as a particular coordinate representation of the original equation (2.3) in an abstract innerproduct space in which random variables are vectors and the covariance function, Cov , forms an innerproduct. We must take the elements of this space to be equivalence classes of random variables defined by $X \sim Y$ iff for all vectors V , $\text{Cov}(X, V) = \text{Cov}(Y, V)$,

otherwise Cov is only a symmetric bilinear form. These elements will be labelled as \underline{X} , \underline{Y} , etc. This is best seen by rewriting these equations in matrix form:

$$\begin{bmatrix} \text{Cov}(Z, X) \\ \text{Cov}(Z, Y) \\ \text{Cov}(Z, Z) \end{bmatrix} = r_{Z/X} \begin{bmatrix} \text{Cov}(X, X) \\ \text{Cov}(X, Y) \\ \text{Cov}(X, Z) \end{bmatrix} + r_{Z/Y} \begin{bmatrix} \text{Cov}(Y, X) \\ \text{Cov}(Y, Y) \\ \text{Cov}(Y, Z) \end{bmatrix}$$

This is the matrix representation of a vector equation

$$\underline{Z} = r_{Z/X} \underline{X} + r_{Z/Y} \underline{Y} \quad \dots, (2.9)$$

relative to the (non-orthogonal) basis $\{\underline{X}, \underline{Y}, \underline{Z}\}$.

In the indeterministic case, $\text{Cov}(E_Z, Z)$ is non-zero (i.e. there are "error" fluctuations) and then it is the (orthogonal) projection of \underline{Z} , rather than \underline{Z} , that is equal to the right hand side of equation (2.9). The other equations, (2.1) and (2.2) can be similarly represented in this abstract innerproduct space.

These relationships are very useful in mathematically representing statistical data, but the contentious question is what role they can perform in causal inference. Usually a functional definition of causal influence is taken for granted as, for instance, X causally influences Z if, and only if, Z is a function of X . But notice that equations (2.6) and (2.7) can be solved uniquely for the unknowns $r_{Z/X}$ and $r_{Z/Y}$ (by Cramer's rule)

if there is an imperfect correlation between X and Y (which is most of the time we can safely assume). But this means that a cyclic causal ordering will follow from the empirical data most of the time if we apply the methods of path analysis to equations (2.1), (2.2), (2.3), and this is unacceptable. So the standard view is that path analysis cannot provide any empirical justification for the causal ordering amongst random variables. Rather, that must be given independently, usually by equating causal order with temporal order. Once this is given, the methods of path analysis are useful for finding the strengths of the causal influences only.

Suppose, then, that the temporal ordering amongst X , Y , and Z is $X \rightarrow Y \rightarrow Z$. This is taken to mean that X cannot be a function of Y or Z , and Y cannot be a function of Z . This reduces the equations (2.1), (2.2), and (2.3) to:

$$Y = r_{Y/X} \cdot X + E_Y \quad \dots (2.10)$$

$$Z = r_{Z/X} \cdot X + r_{Z/Y} \cdot Y + E_Z \quad \dots (2.11)$$

With this model it is more reasonable to take the functional definition of causality more seriously and we can allow the model to dictate what inferences can be drawn from the statistical data about the strengths of the causal connections.

2.5 Bunzl's Theory of Causal Priority

Bunzl starts with something equivalent to equations (2.1), (2.2), and (2.3). Bunzl, however, is unhappy about equating the causal ordering, or priority, with temporal ordering, for this makes backwards causation impossible a priori.

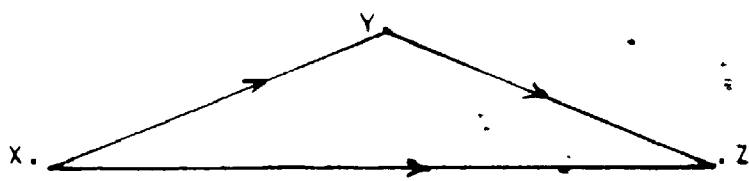
He sets to work, therefore, to find a set of conditions that, together with empirical data (i.e. a set of correlation coefficients among the variables X, Y, and Z) suffice to determine a unique causal ordering, without having to invoke temporal assumptions. His assumptions are:

(1) The "error terms" E_x , E_y , and E_z are mutually uncorrelated, i.e. $\text{Cov}(E_x, E_y) = \text{Cov}(E_y, E_z) = \text{Cov}(E_x, E_z) = 0$.

(2) The assumption of causal priority: For any pair $r_{x/y}$ and $r_{y/x}$, one can be set to zero, though we do not know which one.

(3) "We shall make the further assumption that the dependencies are linear" [p.37]

By assumption (3) Bunzl means to specify that any two variables are only connected linearly in the sense that either there is a direct causal connection between them or an indirect one, but not both. So by assumption (3) he eliminates possible causal orderings as:



What is the justification for this assumption? Where did it come from? Bunzl gives no explanation. Assumption (1), on the other hand, is motivated by the idea that error terms should be uncorrelated because they come from outside the system. The need for assumption (2) is obvious enough, but Bunzl fails to explain explain exactly why causal asymmetry is related to the equation asymmetries as reflected by the matrix of coefficients. If I've ever seen a bunch of ad hoc assumptions, this is them.

But worse than that they don't even do a very good job! First, as Bunzl admits, there are cases where the empirical data fail to determine a unique causal ordering under these three assumptions. The deterministic examples, (a) and (b) above, are sufficient to make this point, because all three of Bunzl's assumptions are satisfied there but the correlations underdetermine the causal order. Secondly, assumption (3) is not only ad hoc, but it will be provably false in many cases. Consider a case in which both atmospheric pressure (X) and humidity (Y) affect the probability of rain (Z), and

pressure (X) affects the humidity (Y) by changing the evaporation rate. The correlations can be such that no linear causal ordering can explain the data.

Could Bunzl do without assumption (3)? Not without some losses because it would then be false that

Where a zero correlation obtains [$\text{Cov}(X,Z)=0$] it is because X and Z are unconnected except insofar as they are the causal antecedents of a third variable.

To show this, suppose the, "nonlinear" causal ordering obtains as diagrammed above, and the coefficients $r_{Y/Z}$, $r_{X/Y}$, and $r_{X/Z}$ are zero. Then we can prove that (see section 2.4)

$$\text{Cov}(X,Z) = [r_{Z/X} + r_{Z/Y} \cdot r_{Y/X}] \text{Cov}(X,X)$$

and it is perfectly possible that $[r_{Z/X} + r_{Z/Y} \cdot r_{Y/X}]$ is zero. In that case X and Z will not be correlated, but they are causally connected, both directly and indirectly. It's just that the sum of those two influences add to give a zero resultant.

Hopefully the reader is by now sufficiently dissatisfied with Bunzl's theory to give the alternative "intervention" theory of causation a fair hearing. There will be two parts to my project. Firstly, in section 1 of chapter 3, I will show that the intervention account of causality can be given the same metaphysical

respectability as in the usual treatments of path analysis in which causal ordering is simply given. On both accounts the measures of causal influence reduce to the constant coefficients in the equations that characterize the system. With respect to the epistemological problems, the intervention account does a lot better. This is treated in last section of the next chapter.

Chapter 3

A RE-INTERPRETATION OF PATH ANALYSIS

3.1 The Deterministic Case

The two causal hypotheses considered for the electrical circuit examples in the last chapter, (a) and (b), can be represented in terms of a network of basic functional dependencies that hold among the random variables. Suppose that a system is modelled by the set of basic equations:¹

$$Y = r_{Y/X} \cdot X + e_Y \quad \dots (3.1)$$

$$Z = r_{Z/X} \cdot X + r_{Z/Y} \cdot Y + e_Z \quad \dots (3.2)$$

where $r_{Y/X}$, $r_{Z/X}$, $r_{Z/Y}$, e_Y and e_Z are given constants. The reason for this particular notation will become clear later. For the 'common cause' case $r_{Z/X} = 0$, and for the 'serial' case $r_{Z/X} = 0$. In general, when none of

1. The important distinction between basic and derived equations will be discussed later in this section. For now, it suffices to think of these equations as "basic" in the sense that they cannot be derived from anything else.

the r-coefficients is zero, we have a mixture of common and serial causation. The basic equations holding among the 'chosen' random variables in the relay switch examples are; for case (a);

$$Y(t_1) = X(t_0) \quad \dots (3.1a)$$

$$Z(t_2) = X(t_0) \quad \dots (3.2a)$$

while for case (b);

$$Y(t_1) = X(t_0) \quad \dots (3.1b)$$

$$Z(t_2) = Y(t_1) \quad \dots (3.2b)$$

Clearly the two models are mathematically equivalent, but they are not to be treated as interpretationally equivalent. The two systems, (a) and (b), do have different properties although this does not show up in the overt behaviour in normal conditions. The overt behaviour of the two systems is isomorphic and hence described by two sets of equations having identical or isomorphic solution sets. Yet we want to say that the systems are of fundamentally different types because they have different causal structures. What marks this difference is the truth of various counterfactual conditionals about what would happen to the system were it interfered, with in certain ways. With respect to these truths the systems (a) and (b) are radically different. But, clearly, if this is to mark a real difference, with the emphasis on

"real", then we must take these counterfactual conditionals to be true or false in virtue of some property possessed by that system.

Information about such causal properties can be encoded in the basic equations, provided we make a principled distinction between basic and derived equations and then follow a set of rules about which equations remain valid under counterfactual conditions in which various constraints are imposed upon the system. The difference between (a) and (b) can then be read from the different sets of basic equations, even though the two sets of equations have the same solution sets. These are the ideas to be developed in this section.

From the equation (3.2), we can use the methods of path analysis to arrive at the same equations as before; namely

$$\text{Cov}(Z, X) = r_{Z/X} \cdot \text{Cov}(X, X) + r_{Z/Y} \cdot \text{Cov}(Y, X)$$

$$\text{Cov}(Z, Y) = r_{Z/X} \cdot \text{Cov}(X, Y) + r_{Z/Y} \cdot \text{Cov}(Y, Y)$$

$$\text{Cov}(Z, Z) = r_{Z/X} \cdot \text{Cov}(X, Z) + r_{Z/Y} \cdot \text{Cov}(Y, Z)$$

The regression coefficient, say, from X to Z will be written $R(Z/X)$, and is defined as

$$R(Z/X) = \text{Cov}(X, Z) / \text{Cov}(X, X)$$

If we now divide through by all the self-covariances, or variances, then these three equations are equivalent to (provided $\text{Cov}(X, X)$, etc., are non-zero)

$$R(Z/X) = r_{z/x} + r_{z/y}.R(Y/X) \quad \dots (3.3)$$

$$R(Z/Y) = r_{z/y} + r_{z/x}.R(X/Y) \quad \dots (3.4)$$

$$1 = r_{z/x}.R(X/Z) + r_{z/y}.R(Y/Z) \quad \dots (3.5)$$

The question now is by what rules do we read off the values of $R(Z/[X])$, $r(Z/[X])$, etc. The idea here is that the counterfactual situation in which the value of Y , say, is imposed is one in which the equation for Y is inoperative. The usual causal ancestry of Y is no longer efficacious. Instead, the value of Y is determined by some different chain of events unconnected with the system (as defined) - that is what is meant by imposing the value of Y from outside the system. This will be summarized by two rules, but first we need to introduce some notation. Denote the original system of equations governing X , Y , and Z by $S(X,Y,Z)$, and denote the set of equations governing the system modified by the external imposition of certain variables by enclosing those variables in square brackets, e.g. $S([X],Y,[Z])$ denotes the set of equations operating when the values of X and Z are imposed from outside. The question is what is the set of equations governing such modified systems?

Rule(1): The modified system of equations, $S(X,[Y],Z)$ say, is obtained from the original set $S(X,Y,Z)$ by deleting the equation governing any variable in square

brackets, i.e. deleting the equation in which that variable is the dependent variable. The equation governing Y is equation (3.1) so that while $S(X,Y,Z) = \{\text{eqn}(3.1), \text{eqn}(3.2)\}$, we have $S(X,[Y],Z) = \{\text{eqn}(3.2)\}$.

Rule(2): (i) Any total causal coefficient is evaluated as a regression coefficient relative to the set of probability functions consistent with the system of equations modified by deleting the equations governing just those variables enclosed in square brackets in the causal coefficient. So $R(Z/[X])$ is evaluated as a regression coefficient $R(Z/X)$ relative to $S([X],Y,Z)$. (ii) Any direct causal coefficient, $r(Z/[X])$ say, is evaluated as a regression coefficient $R(Z/X)$ relative to the system of equations where all equations are deleted except the one governing the variable not in square brackets (Z in this case). Thus $r(Z/[X])$ is evaluated as $R(Z/X)$ relative to $S([X],[Y],Z)$. Partial regression coefficients are no longer used. It is also assumed that any two variables that are not functionally related are stochastically independent. This assumption plays a major role in evaluating the causal coefficients.

For the particular system, $S(X,Y,Z) = \{\text{eqn}(3.1), \text{eqn}(3.2)\}$, that we have been considering, imposing the value of X makes no difference, because there is no equation governing X . So let us consider $R(Z/[Y])$ as

a non trivial example. We consider therefore $S(X,[Y],Z) = \{eqn(3.2)\}$ and assume X and Y to be stochastically independent, as they are not functionally connected in $S(X,[Y],Z)$. This assumption is reasonable in that then we would expect any interference in the value of Y to randomize the connection between Y and its usual causal antecedents (X in this case). On the assumption that X and Y are stochastically independent (which implies that $Cov(X,Y) = 0$) we have that $R(Y/X) = 0 = R(X/Y)$ so that equations (3.3) and (3.4) tell us that

$$r(Z/[X]) = r_{Z/X} \quad \dots (3.6)$$

$$r(Z/[Y]) = r_{Z/Y} \quad \dots (3.7)$$

The major result here is that the direct causal coefficients are identical to the equation coefficients in the basic (linear) equations governing the system, in this case $\{eqn(3.1), eqn(3.2)\}$. In this way the notion of causality, as measured by these coefficients, receives high marks in cross-situational invariance because these constants are involved in explaining how the system behaves in a variety of different situations.

2. But the converse is not true. This fact is noted in any good textbook on probability theory, e.g. Feller (1968). One counterexample is obtained by supposing that $Z = \sin(X)$ and $Y = \cos(X)$, and taking a equiprobable distribution for X over the domain $(0, 2\pi)$. We may verify that $Cov(Y,Z) = 0$, but Y and Z are clearly not independent. (Beckman, 1967, p.88).

To summarize, if Z is a linear function of X and Y with constant coefficients, as in (3.2), then the coefficient $r_{z/x}$ is equal to the direct causal coefficient $r(Z/[X])$. Similarly the equation coefficient $r_{z/y}$ is the causal coefficient $r(Z/[Y])$, and $r_{y/x}$ is equal to $r(Y/[X])$. Note that this inference is valid for any linear network of equations, including the case where there are cyclic loops. (Such a two-way causal model will be used in chapter 5.)

Note that the derivation does not depend on the random variables X , Y , and Z being two-valued.

For the special case where we have dichotomous random variables, and where X and Z take on the values v^+ or v^- , the regression coefficients are simply expressed in terms of probabilities as

$$R(Z/X) = [P(Z=v^+/X=v^+) - P(Z=v^+/X=v^-)] \quad \dots (3.8)$$

The proof of (3.8) proceeds from the definitions of $\text{Cov}(X, Z)$ and $\text{Var}(X)$.

$$\begin{aligned} \text{Cov}(X, Z) &= \langle X, Z \rangle - \langle X \rangle \langle Z \rangle \\ &= (v^+)^2 [P(X=v^+, Z=v^+) - P(X=v^+)P(Z=v^+)] \\ &\quad + v^+v^- [P(X=v^+, Z=v^-) - P(X=v^+)P(Z=v^-)] \\ &\quad + v^-v^+ [P(X=v^-, Z=v^+) - P(X=v^-)P(Z=v^+)] \\ &\quad + (v^-)^2 [P(X=v^-, Z=v^-) - P(X=v^-)P(Z=v^-)] \end{aligned}$$

$$= (v^{+2} - 2v^{+}v^{-} + v^{-2})x \\ [P(X=v^{+}, Z=v^{+}) - P(X=v^{+})P(Z=v^{+})]$$

$$= (v^{+} - v^{-})^2 [P(X=v^{+}, Z=v^{+}) - P(X=v^{+})P(Z=v^{+})]$$

From here it also follows that

$$\text{Var}(X) = (v^{+} - v^{-})^2 [P(X=v^{+}) - P(X=v^{+})P(X=v^{+})] \\ = (v^{+} - v^{-})^2 [P(X=v^{+})P(X=v^{-})]$$

Dividing these two expressions gives

$$R(Z/X) = \frac{P(Z=v^{+}/X=v^{+}) - P(Z=v^{+}, X=v^{+}) - P(Z=v^{+}, X=v^{-})}{P(X=v^{-})} \\ = ([1 - P(X=v^{+})]/P(X=v^{-})) [P(Z=v^{+}/X=v^{+}) \\ - P(Z=v^{+}/X=v^{-})] \\ = [P(Z=v^{+}/X=v^{+}) - P(Z=v^{+}/X=v^{-})]$$

The probabilistic component of expression (3.8) has the same qualitative properties as the expression

$$P(Z=v^{+}/X=v^{+}) - P(Z=v^{+}) \quad \dots (3.9)$$

Both coincide at the extreme values -1, 0, and +1, and both always have the same sign. The event X has a causal influence on Z just in case the expression given in (3.9) is strictly greater than zero, where P is consistent with a certain modified set of equations. The intuitive idea behind such a measure is that a (positive) cause must increase the probability of its effect and that under

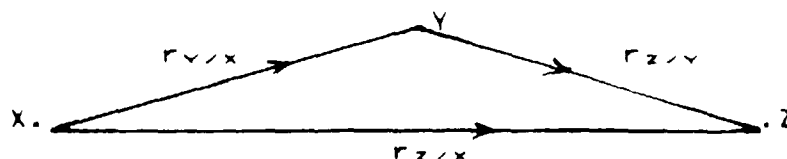
certain counterfactual conditions, namely those in which the antecedent event is imposed from outside the system, the total increase in probability can only be attributed to that cause. The rules given above constrain what is meant by that.

It is easy to check that we get the right answers for the relay switch examples with which we started. By looking at equations (3.1a) and (3.2a) we see immediately that $R(Z/[Y]) = r(Z/[Y]) = 0$, so there is no causal connection from Y to Z, as we would expect. But in case (b) there is such a connection, and indeed $R(Z/[Y]) = r(Z/[Y]) = 1 > 0$. In both cases there is no backwards causation as $r(X/[Z]) = r(Y/[Z]) = r(X/[Y]) = 0$.

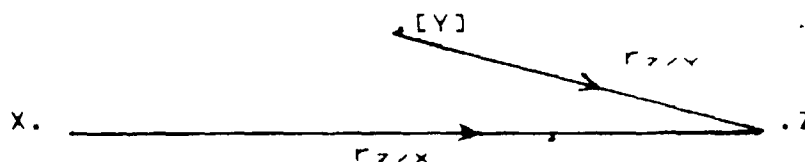
Of course what does the work in determining the causal structure is the structure inherent in the network of basic equations. This structure can be represented diagrammatically in terms of directed graphs, where the equation coefficients are associated with the directed edges. For instance, the weights of the edges for the graph of {eqn(3.1), eqn(3.2)} are given as the coefficients appearing in those equations - and conversely the structure of those equations can easily be read off the graph.³ The graph for {eqn(3.1), eqn(3.2)} is drawn

3. See, for instance, Abrahams, et al. (1965).

below.



Suppose that we now consider the counterfactual situation in which Y is imposed on the system from outside. Following rule (1) the graph of resulting set of equations is drawn below:



All edges directed into Y have been erased, signifying that those causal pathways have been blocked by intervention.

In such a situation, the ordinary regression coefficients correspond to the coefficients in the equation, as we proved earlier. E.g. $r_{Z/X} = R(Z/X)$ in the counterfactual situation indicated by the above graph. In the case of dichotomous random variables this reduces to the definition of causal influence given in the last section.

Return, now, to the normal situation in which both

(3.1) and (3.2) are operative. Here the manifest regression from X to Z is a more complicated function of the parameters of the model:

$$R(Z/X) = r_{z/x} + r_{y/x} \cdot r_{z/y} \quad \dots (3.10)$$

which is derived by substituting $Y=r_{y/x} \cdot X$ into $Z=r_{z/x} \cdot X + r_{z/y} \cdot Y$, to obtain Z as solely a function of X. These relationships, such as (3.10), are usually obtained by a set of rules that are known as Wright's rules, which read off such relationships from the associated directed graph.⁴ Wright's rules are a special case of Mason's rules, the latter are more general in that they apply also to equation systems with cyclic loops.⁵ Equation (3.10) gives the manifest correlation between X and Z in normal circumstances, as measured by the regression coefficient $R(Z/X)$, as a function of direct causal coefficients.

This interpretation of equation coefficients as measuring direct causal influence applies only to basic equations, such as equation (3.1) or equation (3.2). Derived equations, in contrast, can not only be changed by interfering with the dependent variables (just as for basic equations) but also by manipulating other

4. See Wright (1968) for a statement of the rules as well as back references.

5. For a statement of Mason's rules see Abrahams (1965, p.31). For a proof see Lorens (1956).

variables. In our particular example we can derive the equation

$$Z = [r_{z/x} + r_{y/x} \cdot r_{z/y}] \cdot X \quad \dots (3.11)$$

by substituting equation (3.1) into (3.2). If we interfere with the naturally determined value of Y - rendering (3.1) inoperative - we will change this functional connection between X and Z (equation (3.11)). But equation (3.2), being basic, will remain intact. Another example is the equation

$$Z = [r_{z/y} + r_{z/x} / r_{y/x}] \cdot Y \quad \dots (3.12)$$

again derived from (3.1) and (3.2). This describes the correlation between Y and Z in normal situations. But again this deterministic relationship is destroyed when the value of Y is determined independently of the natural evolution of Y. Basic equations can be rendered inoperative too but they are more robust than derived equations in that they can only be changed by manipulating the dependent variable.

Only basic equations thereby have the special status of reflecting direct causal relationships, as opposed to correlations arising indirectly from causal chains or from "spurious" common causes. It is only direct causes that are accorded the proper ontological status, all derived functional dependencies are explained in terms of them.

Of course, whether a causal relation is taken to be direct is relative to the specification of the system. As more variables are added to the model a causal relation that was previously "direct" will be analysed as arising through a chain of intervening variables. The attribution of "basic ontological status" is therefore revisable in light of a more complete description of reality. And what marks that difference in ontological status is the "robustness" of the equation from which the regression coefficient is read - "robustness" in the sense of resistance to change when variables - not mentioned in the equation but included as a system variable - are controlled and manipulated from outside the system.⁶

The successful mathematization of nature is, therefore, support for this realist understanding of causality. Questions concerning our knowledge of causal laws reduce to questions about our epistemic access to the basic equations governing natural phenomena. The search for robust regularities forms the basis of the reductionist strategy in science. This view of causation also accords with standard experimental method. In our original

6. I believe that the principle of realism formulated in this way is connected with the intent behind Wimsatt's (1980) discussion of scientific realism in terms of the principle he calls "robustness as a criterion of reality". Wimsatt does not allude to the concept of "intervention", but Hacking (1982,1983) does in formulating his version realism.

example, what experiment would test between hypotheses (a) and (b)? The crucial experiment will require that we impose the condition $Y=-1$ on TV₁ (a bright screen) in cases when $X=+1$ and imposing $Y=+1$ (a dark screen) when $X=-1$. If the correlation between Y and Z persists in such cases, i.e. is robust, then this is good evidence that the common cause hypothesis is false and (b) is the case. In the relay switch examples similar tests could be made by manipulating the switches on or off manually at various times to see what happens.

The view that causal laws can only be founded upon passively observed regularities (if at all) is represented by the use of partial regression coefficients such as $R_{Y(Z/X)}$ as measures of causality. The standard empiricist arguments then lead to the skeptical conclusion that we can never be justified in believing that causal connections really exist "out there" because all we can observe are correlations and these underdetermine the causal facts. Causal inferences are based merely on psychological habits that we learn when we are young or are maybe even innate. Let me say here that I have some sympathy with this latter view in that I believe that the role that reason plays in our belief formation is over-rated by some philosophers. But what I do object to is any sharp division being drawn between so-called extensional predicates such as "being red" and modal

predicates such as "being caused by X". "Being red" is a nice example because the correct attribution of this predicate also involves counterfactual suppositions, e.g. "X is red" implies that "it would look red if it were viewed in sunlight". Similarly "X is a cause of Y" is taken here to imply that "if the correlation between X and Y were to be viewed in situations in which the value of X were imposed from outside then the correlation would be non-zero⁷". The point worth noticing here is that even our ability to know that something is red is contingent upon our ability to actively control and adapt to the conditions of perception, rather than just some mysterious ability to (passively) observe the world. Conversely, of course, our interest in knowing the world is to actively change things in desirable ways. It is my hope that this may open the way for a deeper view of these matters, but no attempt has been made to work that out here. More will be said about the epistemological aspects of causality in section 4.

The aim of this section has been more modest, namely to explicitly state various rules for inferring causal and other modal facts from the robust basic equations that

7. This way of viewing epistemology - with emphasis on the active two-way interaction between us and the external environment - has been developed at length in Hahlweg (1983) as a new approach to evolutionary epistemology.

characterize a system. Of course we need strong premises to support strong conclusions, and indeed there is a lot built into the assumption that a given set of equations is basic. But there are also assumptions about what it means for an event to be imposed from outside the system as well as the restriction that the equations be linear. So the analysis is not simply a matter of building the desired causal structure into the basic equations with which we started, and therefore question begging. Of course, this brief remark is not a complete defense of my analysis by any means. The strength of the analysis, rather, lies in the way a number of different philosophical issues can be threaded together to provide an interesting account of scientific practice, scientific laws, counterfactual reasoning, and causal inference.

3.2 Generalization to Stochastic Models

The focus on deterministic systems in the last section served its purpose, but clearly I need to show that the view of causality developed there can compete with the traditional account in strictly indeterministic cases.

Let me state the theorem central to this section for the restricted case of just 3 random variables X , Y , and Z . It will be obvious how the theorem is generalized.

THEOREM 1:

Suppose that a system is described by the basic equations

$$\langle Y \rangle_x = r_{y/x} \cdot X + e_y \quad \dots (3.13)$$

$$\langle Z \rangle_{xy} = r_{z/x} \cdot X + r_{z/y} \cdot Y + e_z \quad \dots (3.14)$$

where all the r 's and e 's are given constants. Suppose that normal conditions hold. Then the following phenomenological regression coefficients are related to the parameters of (3.13) & (3.14) just as for the deterministic case;

$$R(Y/X) = r_{y/x} \quad \dots (3.15)$$

$$R(Z/X) = r_{z/x} + r_{z/y} \cdot R(Y/X) \quad \dots (3.16)$$

$$R(Z/Y) = r_{z/y} + R(X/Y) \cdot r_{z/x} \quad \dots (3.17)$$

The same arguments as before then show that the direct causal coefficients can again be equated to the corresponding equation coefficients. That is, $r(Y/[X]) = r_{y/x}$, $r(Z/[X]) = r_{z/x}$, and $r(Z/[Y]) = r_{z/y}$.

PROOF OF THEOREM 1: We will only prove that $R(Y/X) = r_{y/x}$ because the rest will follow analogously to the deterministic case. Take equation (3.13). Consider probability measures satisfying that equation such that

$\text{Var}(X) = 0$, i.e. where a certain value of X , say x , is assigned probability 1. The value of $\langle Y \rangle$ for that probability function has been denoted by $\langle Y \rangle_x$. Then

$$\langle X \langle Y \rangle_x \rangle = \sum_x x \langle Y \rangle_x P(X=x) = \sum_x x \sum_y y \cdot P(Y=y/X=x) P(X=x) = \langle X \cdot Y \rangle$$

From (13) we know that $\langle Y \rangle_x \cdot x = r_{Y/X} \cdot x^2$ so that $\langle X \cdot Y \rangle = r_{Y/X} \cdot \langle X^2 \rangle$. Its now a short step to the desired result. The important premise for the proof is that

$$\langle Y \rangle_x = \sum_y y \cdot P(Y=y/X=x)$$

where $P(* / X=x)$ is obtained from $P(*)$ by conditionalizing on the proposition $X=x$.

It should be emphasized here that the constants in equations (3.13) & (3.14) are not intended to represent the statistical properties of some fictitious ensemble of systems, but rather the actual property of the single system. It is just that this property is most conveniently referred to in terms of a "counterfactual locution" that exploits various statistical concepts, such as relative frequency over an ensemble. However, the ontological status of such properties need be no different from more familiar observational properties (such as being red) on my view. The fact that the physical property of having a certain surface reflectance can be referred to as the property of "appearing red had it been viewed under sunlight" does not mean that surface reflectance is a

property of some fictitious object being looked at in sunlight. Thus the probabilistic properties described by (3.13) and (3.14) apply to the single system. The possible worlds account of how to understand counterfactual locution is misleading on exactly this point, because it takes the truth conditions of counterfactual conditionals as defined in terms of what happens in a statistical ensemble of copies of the system or 'nearest possible worlds'. It is true that the only verification procedures that we can have involve an ensemble of identical systems, but I do not wish to conflate truth conditions with verification conditions. For what we seek is, in part, an explanation of why such an ensemble behaves as it does, and this involves positing certain properties to each system in the collection - each system is after all assumed to be physically isolated from the other and the ensemble has no properties other than those of its parts. So if the counterfactual conditional is true, then it is true in virtue of causal properties possessed by the individual system - otherwise the regularities inherent in the behaviour of the ensemble would remain a mystery.

3.3 What Are the Causal Relata? Events or Universals?

It is appropriate here to say more about the move from dichotomous (2-valued) to many-valued or even continuous random variables. Earlier we discussed the reduction of the regression coefficient to the probabilistic expression (3.8) in the case of dichotomous random variables each of which takes on the same two values (v^+ or v^-). It is expression (3.8) that captures the intuition that a cause should increase the probability of its effect. But here a 'cause' is clearly meant to be an event. I want to argue here that this common sense intuition should be abandoned because it does not generalize to the more general case of non-dichotomous or multi-valued random variables.

For the general dichotomous case, where X equals x^+ or x^- but Z equals z^+ or z^- , we find that

$$R(Z/X) = (z^+ - z^-) / (x^+ - x^-) [P(Z=z^+/X=x^+) - P(Z=z^+/X=x^-)] \quad \dots (3.18)$$

The dimensionless factor $[P(Z=z^+/X=x^+) - P(Z=z^+/X=x^-)]$ is symbolized $Q(z^+/x^+)$. The dimensional factor indicates how much of a difference in X can possibly make to Z - the numerical value assigned here clearly depends on the units

or the scale chosen for the variables X and Z . When we change the measurement scale for Z , say, by a factor k then $R(Z/X)$ increases by a factor k since

$$R(Z/X) = [\langle Z \rangle' - \langle Z \rangle''] / [\langle X \rangle' - \langle X \rangle'']$$

by the result to be proved in the section 3.5, although it is clearly invariant with respect to translational transformations of the origin. But $Q(z^*/x^*)$ does not change - it is invariant under scale changes. The question to be addressed here is whether this feature of regression coefficients - the factorizability into a dimensional and a non-dimensional component - generalizes to the case of multi-valued random variables.

There is an expression for the regression coefficient R that factors into a dimensional and a dimensionless factor.

$$R(Z/X) = \frac{\sigma(Z)}{\sigma(X)} p(X,Z) \quad \dots (3.19)$$

where ' σ ' symbolizes the 'standard deviation' and is the square root of the variance, and $p(X,Z)$ is the correlation coefficient. In terms the covariance,

$$p(X,Z) = \text{Cov}(X,Z) / [\sigma(X) \cdot \sigma(Z)] = \text{Cov}(X/\sigma(X), Z/\sigma(Z)).$$

The disadvantage of $p(X,Z)$ is that the robustness of $R(Z/X)$ does not guarantee the robustness of $p(X,Z)$. On

the other hand, the robustness of $Q(z^+/x^+)$ does follow from the robustness of $R(Z/X)$ in the dichotomous case because the factor $[z^+-z^-]/[x^+-x^-]$ is a constant (does not depend on which probability distribution is used to evaluate $R(Z/X)$, unlike the analogous factor in equation (3.19)). But even if $p(X,Z)$ were guaranteed invariant there is still a significant difference. $Q(z^+/x^+)$ can be said to be a relation between two events, whereas $p(X,Z)$ is a relation between two random variables. The common sense notion of cause, as I take it, is considered to be a relation between two events. It is interesting to ask, therefore, whether there is anything corresponding to $Q(z^+/x^+)$ in the general case. Unfortunately $Q(z^+/x^+)$ is only defined in an obvious way for dichotomous random variables. So much the worse for common sense.

What I propose, therefore, is that the relata of the causal relata be random variables. To make sense of this move will require some metaphysical rethinking. But that is well overdue on independent grounds anyway. One possible approach would be to associate appropriate random variables with universals; after all it is essential to the epistemological aspects of 'cause' that the phenomena be repeatable, that we examine the behaviour of copies of the same system (whatever that's taken to mean). Universals are exactly characterized as 'repeatable entities' [Lewis, 1983, p.343], and so viewing the causal

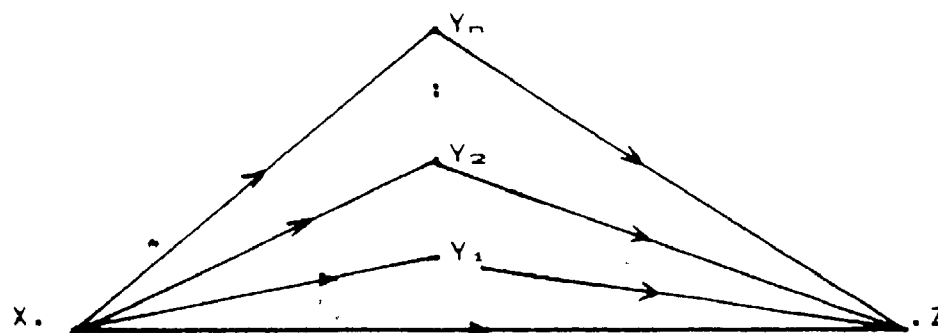
relation as relating two universals would fulfill this role, as well as connections with the Dretske-Tooley-Armstrong view of theories. It looks like this may be more "new work for a theory of universals" [Lewis, 1983].

But the crucial argument to be made here takes advantage of the discussion of cross-situational invariance as presented in chapter 1. Taking the causal relata to be events would damage the usefulness of the causality concept in capturing what is common to the system in a wide variety of situations, at least for the class of models considered here. For $Q(Z=z/X=x)$, for some values z and x of Z and X , is not connected in any way to the parameters of these models - whereas $R(Z/[X])$, etc., does have this crucial property. So $R(Z/[X])$, etc., is the theoretical construct that can be taken seriously on the basis of a cross-situational invariance, and this is a relation between random variables, not events.

3.4 The Question of Transitivity.

Wright's rules give a simpler and more perspicuous representation of the "transitivity theorem" proved in Eells & Sober (1983). To state that theorem, we first need to generalize the example we have been considering by

replacing the random variable Y with $Y_1, Y_2, Y_3,$ etc., with possible causal connections as indicated by the directed graph drawn below.



As is easily seen from the graph, the system of basic equations generalizes to

$$\text{For } 1 \leq k \leq n, \quad \langle Y_k \rangle_x = r_{Y_k/x} \cdot X \quad \dots (3.20)$$

$$\langle Z \rangle_{Y_1 \dots Y_n} = \sum_k r_{Z/Y_k} \cdot Y_k \quad \dots (3.21)$$

From the results proved in earlier sections, it is easy to see that

$$R(Z/[X]) = r_{Z/x} + \sum_j r_{Y_j/x} \cdot r_{Z/Y_j} \quad \dots (3.22)$$

What this says is that the total causal influence from X to Z is equal to the direct causal influence from X to Z plus the sum of all influences acting via an intervening variable each of which is a product of causal coefficients as indicated. Purely within the context of probability theory Eells & Sober prove that if we assume the Markov property, and the mutual independence of the Y 's

conditional upon X , then the total causal influence of X on Z is positive if X has a positive causal influence on all the Y 's and all the Y 's have a positive causal influence on Z . This fact follows easily from (3.22). Under the Markov condition $r_{Z/X} = 0$. So obviously $R(Z/X)$ is positive if $r_{Y_j/X}$ and r_{Z/Y_j} are positive for all j . Of course they use the partial regression analysis, which is only valid in strictly indeterministic cases and when the causal ordering is given (see chapter 2).

Many other qualitative facts not mentioned by Eells & Sober also follow from (3.18), e.g. that if all the r 's are negative then $R(Z/[X])$ is also positive. Besides that (3.18) gives a quantitative and perspicuous representation of such transitivity properties, which extends beyond the special case of dichotomous random variables.

3.5 Epistemological Aspects of the Intervention Account

There is a converse to theorem 1 which is a little suprising, and (as far I'm aware) original. Consider a system with variables identified as X , Y and Z , and labelled as $S(X,Y,Z)$.

Let us start by assuming that we have some way of

identifying "copies" of this system as being of the same type, i.e. as obeying the same set of basic equations. Only that these basic equations are, as yet, unknown. We are now instructed to gather statistical information about how this system behaves under various forms of intervention. Needless to say, I am also assuming that we are in fact manipulating or intervening in the system rather than the system manipulating us (Bunzl, 1984, 1980).

So, in particular, suppose we know the regularities that hold on the modified systems $S([X],[Y],Z)$, $S([X],Y,[Z])$, and $S(X,[Y],[Z])$. What inferences can be made concerning the basic equations governing $S(X,Y,Z)$? This I take to be the epistemological problem of making causal inferences on the basis of statistical data.

This particular question will be answered on the basis of the theorem proved below. The theorem is actually quite general - it will give us a complete characterization of the equations (basic or derived) for any modified system. But the only equation true in $S([X],[Y],Z)$ will be the basic equation for Z , and only the basic equation for Y will be true in $S([X],Y,[Z])$, and only that for X will be true in $S(X,[Y],[Z])$. So the statistical regularities that hold in those three modified systems will completely determine the basic equations in

$S(X,Y,Z)$. Mathematically speaking, the total "statistical information" about a system - characterized by some set of equations - will be represented as the set of all probability functions consistent with those equations. The theorem required will be stated after a preliminary definition.

DEFINITION 1: Any regression coefficient $R(Z/X)$, say, is robust with respect to a set of probability functions consistent with a system S^* iff $R(Z/X)$ takes the same value for all P in S^* , i.e. is invariant with respect to S^* .

Again we assume that any P being consistent with S^* to imply that any two variables not functionally related in any way in S^* will be treated as stochastically independent. In the context of $S(X,[Y],Z)$, say, this is motivated by the idea that the act of imposing the value of a variable Y will randomize it from any normal causal antecedents. So Y will be stochastically independent of X , say, in $S(X,[Y],Z)$ because X is not a function of Y (and Y is no longer a function of anything (in the system) so Y is not a function of X).

The theorem can now be stated as follows:

THEOREM 2: If the regression coefficient $R^*(Z/X)$ is robust with respect to S^* and therefore identifiable with a

constant r^* say, then an equation valid for $\langle Z \rangle$ in S^* is of the form

$$\langle Z \rangle = r^* \langle X \rangle + (\text{other terms not involving } X).$$

PROOF OF THEOREM 2:

Suppose that P' and P'' are both probability functions consistent with S^* that share the same distribution for Y so that $\langle Y \rangle' = \langle Y \rangle''$ but $\langle X \rangle'$ may not be equal to $\langle X \rangle''$. Assume that $\text{Var}(X)'$ and $\text{Var}(X)''$ are non-zero. Let $P = p.P' + q.P''$, where p and q lie between 0 and 1, and such that $p+q=1$. Then P is also a probability function and is also consistent with S^* . First we need the following lemma:

LEMMA: Evaluated in terms of the function $P = p.P' + q.P''$,

$$\text{Cov}(X, Z) = p.\text{Cov}(X, Z)' + q.\text{Cov}(X, Z)'' + pq[\langle X \rangle' - \langle X \rangle''][\langle Z \rangle' - \langle Z \rangle'']$$

PROOF: First note that $\langle * \rangle = p.\langle * \rangle' + q.\langle * \rangle''$, where the $*$ can be replaced with any function. Now by definition,

$$\text{Cov}(X, Z) = \langle [X - \langle X \rangle][Z - \langle Z \rangle] \rangle$$

8. This assumption is to ensure that any two variables that are stochastically independent with respect to P' and P'' remain so with respect to $p.P' + q.P''$.

$$\begin{aligned}
&= p.\langle (X - [p\langle X \rangle' + q\langle X \rangle'']) (Z - [p\langle Z \rangle' + q\langle Z \rangle'']) \rangle' \\
&\quad + q.\langle (X - [p\langle X \rangle' + q\langle X \rangle'']) (Z - [p\langle Z \rangle' + q\langle Z \rangle'']) \rangle'' \\
&= p.\langle (X - \langle X \rangle' - q[\langle X \rangle' - \langle X \rangle'']) (Z - \langle Z \rangle' - q[\langle Z \rangle' - \langle Z \rangle'']) \rangle' \\
&\quad + q.\langle (X - \langle X \rangle'' + p[\langle X \rangle' - \langle X \rangle'']) (Z - \langle Z \rangle'' + p[\langle Z \rangle' - \langle Z \rangle'']) \rangle'' \\
&= p.\langle \text{Cov}(X, Z) \rangle' + q^2 [\langle X \rangle' - \langle X \rangle''] [\langle Z \rangle' - \langle Z \rangle''] \\
&\quad + q.\langle \text{Cov}(X, Z) \rangle'' + p^2 [\langle X \rangle' - \langle X \rangle''] [\langle Z \rangle' - \langle Z \rangle''] \\
&= p.\text{Cov}(X, Z) \rangle' + q.\text{Cov}(X, Z) \rangle'' \\
&\quad + pq. [\langle X \rangle' - \langle X \rangle''] [\langle Z \rangle' - \langle Z \rangle'']
\end{aligned}$$

As an immediate corollary to this lemma, we have

$$\text{Var}(X) = p.\text{Var}(X) \rangle' + q.\text{Var}(X) \rangle'' + pq. [\langle X \rangle' - \langle X \rangle'']^2$$

In order to prove the main theorem, we will use the following arithmetical fact without proof: Let a, b, c, d, e, f, g , and h be any numbers. THEN,

$$\text{if } \frac{a}{e} = \frac{b+c+d}{f+g+h} \quad \text{and} \quad \frac{a}{e} = \frac{b}{f} \quad \text{and} \quad \frac{a}{e} = \frac{c}{g} \quad \text{then} \quad \frac{a}{e} = \frac{d}{h}.$$

From the lemma we know that

$$\frac{\text{Cov}(X, Z)}{\text{Var}(X)} = \frac{p\text{Cov}(X, Z) \rangle' + q\text{Cov}(X, Z) \rangle'' + pq[\langle X \rangle' - \langle X \rangle''] [\langle Z \rangle' - \langle Z \rangle'']}{p.\text{Var}(X) \rangle' + q.\text{Var}(X) \rangle'' + pq[\langle X \rangle' - \langle X \rangle''] [\langle X \rangle' - \langle X \rangle'']}$$

By the assumption of robustness

$$\frac{\text{Cov}(X, Z)}{\text{Var}(X)} = \frac{\text{Cov}(X, Z) \rangle'}{\text{Var}(X) \rangle'} = \frac{\text{Cov}(X, Z) \rangle''}{\text{Var}(X) \rangle''}$$

Therefore our arithmetical fact tells us that

$$\frac{\text{Cov}(X,Z)}{\text{Var}(X)} = \frac{\text{pq.}[\langle X \rangle' - \langle X \rangle"][\langle Z \rangle' - \langle Z \rangle"]}{\text{pq.}[\langle X \rangle' - \langle X \rangle"][\langle X \rangle' - \langle X \rangle"]}$$

$$\text{i.e. } R(Z/X) = \frac{\langle Z \rangle' - \langle Z \rangle"}{\langle X \rangle' - \langle X \rangle"} = r^*$$

where r^* is a constant. Thus

$$\langle Z \rangle' - r^* \langle X \rangle' = \langle Z \rangle" - r^* \langle X \rangle"$$

provided $[\langle X \rangle' - \langle X \rangle"]$ is non-zero.

That is, on the assumption that $R(Z/X)$ is invariant for all probability measures consistent with S^* , it follows that the expression $[\langle Z \rangle - r^* \langle X \rangle]$ is invariant over that same set of probability measures. Then we know that the equation

$$\langle Z \rangle = r^* \langle X \rangle + E_z$$

is valid for S^* , where E_z is a random variable or some function of the expected values of random variables. E_z must be stochastically independent of X , for otherwise we would have

$$\text{Cov}(X,Z) = r^* \text{Var}(X) + \text{Cov}(X,E_z)$$

where $\text{Cov}(X,E_z)$ is non-zero, and so $R(Z/X)$ would not be equal invariantly to r^* . Therefore

$$\langle Z \rangle = r^* \langle X \rangle + (\text{other terms independent of } X)$$

is valid for the situation corresponding to S^* , as we set out to prove.

To understand the theorem it's best to see how it can be applied. Suppose we are faced with the situation as described by equations (3.13) and (3.14) except we don't know these equations. Rather, we are given that the regression coefficient $R(Z/X)$ is robust and equal to r^* in normal circumstances. The theorem tells us that $\langle Z \rangle = r^* \cdot \langle X \rangle + E_z$ in normal situations, which corresponds to the equation derivable from (3.13) and (3.14) where

$$r^* = r_{z/x} + r_{z/y} \cdot r_{y/x}$$

Secondly, suppose that $R(Z/X)$ is also robust within $S([X],[Y],Z)$ which represents the situation in which the values of x and y is imposed from outside. BUT $R(Z/X)$ will not have the same value. This time $R(Z/X) = r_{z/x}$ and by the theorem just proved, $\langle Z \rangle = r_{z/x} \cdot \langle X \rangle +$ (other terms not involving X). If we apply the theorem to both $R(Z/X)$ and $R(Z/Y)$ within $S([X],[Y],Z)$, representing the situation in which the values of both X and Y are determined independently from outside the system, then we will recover the basic equation;

$$\langle Z \rangle_{xy} = r_{z/x} \cdot X + r_{z/y} \cdot Y$$

$$+ \text{(other terms not involving } X \text{ or } Y) \quad \dots (3.23)$$

Because this equation holds in $S([X],[Y],Z)$ it must be basic, and therefore also holds in $S(X,Y,Z)$.

More general inferences are possible. Suppose $R(Z/X)$ is not robust for $S([X],[Y],Z)$ but is robust in $S([X],[Y=y],Z)$, for each y , where the set of equations $S([X],[Y],Z)$ is obtained from $S(X,Y,Z)$ by deleting the equations governing X and Y and substituting y for Y in the remaining equations. That is $R(Z/X) = r^*(y)$ for each P consistent with $S([X],[Y=y],Z)$. Then from the theorem

$$\langle Z \rangle_{xv} = r^*(Y) \cdot X + (\text{other terms})$$

holds in $S([X],[Y],Z)$ and therefore in $S(X,Y,Z)$.

The last paragraph indicates that the assumption of linearity is not essential to the inference schema described here. So, in summary, what the theorem achieves is as follows: GIVEN the assumptions that (1) we can identify different systems as being instances of the same type (e.g. by identifying the circuit connections in the relay switch examples), (2) that we collect statistical data from sufficiently large numbers of these systems so that we can be sure that the correlations exhibited are not due to accidental fluctuations, (3) that we assume that we can observe what events are actualized in each instance without disturbing it, (4) that uncontrolled exogenous factors are not interfering with variables

operating in the system in a way that is not understood, and (5) that we can collect data about the behaviour of such systems under suitable kinds of intervention, THEN we can infer what basic equations hold for the system.

Notice that the first four assumptions must also be assumed by any statistical account of causal inference. So the only additional assumption required for the intervention account is an "intervention" assumption (5). And that one assumption saves the present account from a number of objections that other probabilistic accounts fall prey to (see section 2.2, chapter 2).

Chapter 4

BELL'S PARADOX AND DETERMINISTIC REALISM

4.1 Introduction

The previous chapter discussed one kind of statistical regularity, and the theoretical model that can be constructed in that case. This first step in the modelling process aims to isolate the cross-situational invariances in the phenomena by extrapolating from one system to another. This lower level stage is still theoretical, as the constant coefficients appearing in the basic equations of the model are theoretical constructs. They help explain how the system behaves in a variety of different situations. The next stage would be to explain these regularities, identified in the constancy of those coefficients across situations, in terms of even higher order invariants over an even broader range of phenomena. In the case of the relay switch examples, this is achieved by the broader theory of electrical circuits.

But the procedure is not always that simple. There are

always assumptions made at each stage of model construction. In the last chapter, these were assumptions about what we are doing when we are interfering with a system in certain ways. Contrary to Bunzl (see the discussion in chapter 2), we do have checks on these assumptions. But these checks are not conclusive in themselves - they serve mainly to narrow the range of possibilities. And as Hacking emphasizes, often we test one model on the basis of a better confirmed model that tells us what we are doing to the system under test. Thus we make use of the better established models of electron behaviour in order to investigate weak neutral currents experimentally.

Francis Bacon, the first and the last philosopher of experiments, knew it well: the experimenter sets out to "twist the lion's tail". Experimentation is interference in the course of nature; "nature under constraint and vexed; that is to say, when by art and the hand of man she is forced out of her natural state, and squeezed and moulded [Bacon]". The experimenter is convinced of the reality of entities some of whose causal properties are sufficiently well understood that they can be used to interfere elsewhere in nature. One is impressed by entities that one can use to test conjectures about other more hypothetical entities. In my example, one is sure of the electrons that are used to investigate weak neutral currents and neutral bosons. [Hacking, 1982, pp.75-6]

So at every stage of model construction we rely on assumptions that, though well entrenched, may be wrong. It may be that the road to better theory is a rocky one.

It may be that further progress towards cross-situational invariance can only be achieved by re-indentifying the lower level regularities on which we base our attempt at a higher level unification. This certainly happened during the Einsteinian revolution at the beginning of this century when the very geometrical foundations of Newtonian physics was overhauled in order to achieve a better unification of mechanics and electrodynamics.

This has become a key issue in the foundations of science. An understanding of the tacit assumptions underlying assumptions of our theories is a major part of the invention and assimilation of better theories. Often the task of identifying and understanding these underlying assumptions is hampered either by the complexity or the abstractness of the mathematical formalism. This is especially so for the case of the quantum domain. The abstractness of the Hilbert space formalism contributes to the difficulty of foundational studies in quantum theory.

One program in the foundations of quantum mechanics has been to isolate a list of empirically plausible axioms from which the abstract Hilbert space structure of the formalism can be deduced. It is not clear, however, that this structure is essential to the modelling of quantum mechanical phenomena. Shimony (1977), for example, argues that much of the Hilbert space structure used in modelling

the spin states of spin-1 particles is redundant in the sense of having no verifiable empirical content. As Shimony concludes, "Doubt is therefore thrown upon the program of recovering the Hilbert space formalism from empirically justified assumptions [Shimony, 1977, p.381]". The attempt to achieve higher order unification on the basis of regularities as theoretically represented in Hilbert space formalism may be fundamentally misguided.

Conversely, the Hilbert space formalism may be overly impoverished. Maybe higher order invariants will only reveal themselves once we add hidden parameters in the lower level models. Initially these will only appear as "fudge factors", but if further constraints can be placed on them they may guide us to further experimentation that successfully provides independent modes of detection for those hidden variables. Remember that this is roughly the story that lead to the acceptance of the atomic hypothesis. This is what the "hidden variable" interpretation of quantum mechanics is all about.

The accepted experimental phenomena themselves, however, place rather powerful constraints on the introduction of a hidden variable representation of quantum regularities. This has been the subject of a series of no-hidden-variable proofs, central to the studies done in

the foundations of quantum mechanics in recent years. The Bell proof is one of the more interesting and accessible of these theorems, and will form the focus of this chapter and the next. It was basically Bell's argument that van Fraassen used as an argument against realism at the beginning of chapter 1. This will give us an opportunity to re-evaluate the issue of realism in modern science.

The aim of sections 2, 3, and 4 is to characterize the general nature of non-classical statistics - roughly those statistics for which the simplest cross-situational extrapolation from one measurement context to another leads to a contradiction. This is the crux of Bell's argument. The conceptual part of the argument is treated in section 4, where sufficient conditions for the extrapolation of experimental results are identified and interpreted. These conditions applied to non-classical statistics lead to a contradiction, so by a reductio ad absurdum inference, one of these assumptions must be false if a hidden variable interpretation of quantum mechanics is to be possible. Because the locality assumption is one of the most vulnerable of these conditions, Bell's argument is often described as a proof against local hidden variables in quantum mechanics.

4.2 Bell-type Inequalities as Consistency Constraints

Bell's theorem has traditionally been formulated within the framework of classical probability theory. Probability itself, however, is a theoretical concept, and so these formulations of the paradox give rise to the initial plausibility that the problem might be resolved by developing a non-classical theory of probability. But this possibility is not as plausible as it first appears because the connection between probability and relative frequencies (as mediated by the laws of large numbers and ergodic theorems) is then thrown out of alignment. Reworking this connection is no easier than the original problem it claims to resolve.

To avoid these pitfalls, it is best to formulate as much of Bell's (1964) argument as possible in terms of finite relative frequencies.¹ This clearly separates the mathematico-logical component of the argument (section 2 and 3) from the conceptual component of the argument (section 5).

1. The idea came from the excellent introductory article by d'Espagnat (1979). I found out later that Stairs (1979) had done something similar in arguing against Fine's interpretation of quantum mechanics.

Take the standard formulation of a spin correlation experiment for electron pairs in the singlet state, where spin measurements may be made in one of three directions in physical space, a , b , or c . Consider any finite sample of n electron pairs indexed by a set of natural numbers J . Make the deterministic assumption that spin values exist in all directions, whether or not they are determined by measurement. Pitowsky's (1982, 1983) model is of this type - the state of an electron is only completely determined by simultaneous spin values for all directions in space. Thus the spin state of electron j might be defined in terms of a function $h_j: S^{(2)} \rightarrow \{0,1\}$, where $S^{(2)}$ is the set of points on the unit sphere, or equivalently the set of all directions in 3-dimensional physical space.

Also assume that, because of the strict negative correlation between values of spin measured in any fixed direction or positive correlation for anti-parallel directions, information about spin in the a -direction, say, can be inferred from a measurement of spin in the a -direction on the second electron of a pair. We can thus obtain experimental information about the simultaneous spin of a particular electron in any two directions, but at most two directions.

Let $h_j(A) = 1$ indicate that the left-hand electron of

the j th electron pair has spin up in the a -direction and $h_j(A) = 0$ indicate that the same electron does not have spin up in the a -direction. $h_j(-C) = 1$ says that the left-hand electron of the j th electron pair has spin down in the c -direction, $h_j(A.B) = 1$ indicates that spin is up in both the a -direction and b -direction, and so on. Similarly $h_{j'}(A) = 1$ etc, applies to the right-hand electron of the j th electron pair. Now define

$$h_{j,j'}(A.B) \stackrel{\text{def}}{=} h_j(A).h_{j'}(B)$$

The assumption of spin conservation tells us that $h_{j,j'}(A.A)=0$, thus justifying the inference that $h_j(A)=1$ iff $h_{j'}(A)=0$. In short, $h_{j,j'}$ is the truth function on propositions applying to electron pair (j,j') . But assuming conservation we can work just with h_j .

Under these assumptions, it makes sense to ask whether² or not the following inequality holds.

$$\begin{aligned} 1/n \sum_{j \in J} h_j(A.B) &\leq 1/n \sum_{j \in J} h_j(A.-C) \\ &+ 1/n \sum_{j \in J} h_j(B.C) \dots (4.1) \end{aligned}$$

That this inequality MUST hold is seen by assuming that

2. This inequality is usually referred to as Wigner's version of Bell's inequality or just Wigner's inequality.

$$h_j(C) + h_j(\neg C) = 1, \quad h_j(A.B) = h_j(A).h_j(B), \text{ etc. } \dots (4.2)$$

These are the usual laws of classical logic which constrain the assignment of truth values to complex sentences constructed from negation \neg and conjunction \cdot .³ The argument is a simple reductio ad absurdum proof. If inequality (4.1) were violated, then for at least one j in J we would have

$$h_j(A.B) > h_j(A.\neg C) + h_j(B.C)$$

otherwise inequality (4.1) would be satisfied. Using the laws of classical logic we easily get

$$h_j(A)h_j(B) > h_j(A)(1-h_j(C)) + h_j(B)h_j(C)$$

This implies that,

$$h_j(A)h_j(B) = 1, \quad h_j(A)(1-h_j(C)) = 0, \quad h_j(B)h_j(C) = 0$$

The last two equations imply that either $h_j(A)$ or $h_j(B)$ is zero, contrary to the first equation. Thus assuming that the inequality (4.1) is violated leads to absurdity. THEREFORE inequality (4.1) can never be violated.

3. These laws entail that a partial ordering amongst equivalence classes of sentences, defined by $X \leq Y$ iff $h_j(X) \leq h_j(Y)$ for all truth functions h_j satisfying the constraints (4.2), forms a Boolean algebra.

The above proof emphasises that the mere logical consistency of propositions applying to single electrons is sufficient to derive the Wigner inequalities. One extreme case where the inequalities appear to be violated is illustrated as follows: Introduce A, B, and C as random variables that take values $A=1$ if $h_j(A)=1$ and $A=-1$ if $h_j(A)=0$, etc. Suppose that we find experimentally that in the context in which we measure spin in the a and b directions we find that $A=B$, in the b and c directions we find that $B=C$, and in the c and a directions we find $C=-A$. If we assume that spin values exist even when not measured and that these hidden variables conform to the same regularities as the measured values, then we may extrapolate those statistics from the appropriate measurement context to the sample J as a whole. But then we can derive the contradiction $A=-A$. The quantum mechanical statistics are less extreme, but if we interpret them as applying to measured and un-measured spin values alike then the above proof shows that any violation of Wigner's inequalities implies that there exists some non-empty set of electrons to which some such extreme statistics do apply. And this is logically impossible, so the quantum mechanical statistics cannot apply to un-measured spin values (they do apply to measured spin values as is experimentally confirmed).

Extreme spin statistics are defined as those in which spin values are always the same or always different for every pair of directions - all the probabilities are zero or one. Extreme non-classical statistics are characterized by the violation of the consistency conditions:

$$c_j(A,B) = c_j(A,C) \cdot c_j(C,B) \quad \dots (4.3)$$

where $c_j(A,B)$ is a kind of correlation or covariance coefficient and is defined by

$$c_j(A,B) = h_j(A=B) - h_j(A \neq B) \quad \dots (4.4)$$

I will call this coefficient the co-equality coefficient. Clearly $c_j(A,B)$ is either 1 or -1, depending on whether $h_j(A) = h_j(B)$ or $h_j(A) \neq h_j(B)$ respectively. The rules for co-equality coefficients such as (4.3) can be derived from a consistency requirement on a network of (linear) functional relationships. To see how that works in our case, first verify that $A = c_j(A,B) \cdot B$, $A = c_j(A,C) \cdot C$, and $C = c_j(C,B) \cdot B$. Now, the last two equations give $A = c_j(A,C) c_j(C,B) \cdot B$ and consistency with the first equation requires that (4.3) holds. So the successful extrapolation of these equations across measurement

4. The above is just one of six different possibilities obtained by permuting A, B, and C, and/or replacing A with -A, etc.

contexts necessitates equation (4.3).

It is useful to see how the Wigner inequalities can be derived from the consistency condition (4.3) because this proof strategy generalizes to the Clauser-Horne inequalities as well as to any Bell-type constraint (see section 4.3). By examining all possible cases, it can be seen that (4.3) is equivalent to the two conditions

$$|c_j(A,C) + c_j(C,B)| = 1 + c_j(A,B)$$

$$|c_j(A,C) - c_j(C,B)| = 1 - c_j(A,B) \quad \dots (4.5)$$

Now take averages and use the triangle inequality for absolute values to arrive at

$$|C(A,C) + C(C,B)| \leq 1 + C(A,B)$$

$$|C(A,C) - C(C,B)| \leq 1 - C(A,B) \quad \dots (4.6)$$

where

$$C(A,B) \stackrel{\text{def}}{=} 1/n \sum_{j \in J} c_j(A,B), \text{ etc.}$$

To see that (4.6) is equivalent to the Wigner inequalities, observe from (4.4) that

$$c_j(A,B) = [h_j(A) - h_j(-A)][h_j(B) - h_j(-B)]$$

so that

$$C(A,B) = P(A,B) - P(A,-B) - P(-A,B) + P(-A,-B)$$

$$= 4P(A,B) - 2P(A) - 2P(B) + 1 \quad \dots(4.7)$$

where $P(A,B) = 1/n \sum_{j \in J} h_j(A)h_j(B)$, etc.⁵

These proofs are easily extended to the Clauser-Horne inequalities⁶ [See Clauser, J.F. & Horne, M.A. (1974) or Fine (1982) for a recent reference on the subject], which can be written as

$$0 \leq 1/n \sum_{j \in J} H_j \leq 1$$

where

$$H_j = h_j(A) + h_j(S) + h_j(B,T) - h_j(A,S)$$

$$-h_j(A,T) - h_j(B,S)$$

where A and S are propositions about the first electron whereas B and T are spin propositions about the second electron of a correlated pair, corresponding to directions a, s, b, and t respectively (see Fig.4.1(b)). To show that this inequality must always be satisfied, apply the conjunction law, and then consider two cases; (i) $h_j(S)=1$, (ii) $h_j(S)=0$. Then divide each of these cases into two

5. Within the framework used by Accardi (1982), inequality (4.5) is equivalent to Accardi's condition (3), which he gives as a necessary and sufficient condition for the existence of hidden variables (a 'Kolmogorovian model' in Accardi's language).

6. I am grateful to Professor Bub who pointed this out to me.

cases; $h_j(T)=1$ and $h_j(T)=0$. For each of the 4 cases, H_j is either 0 or 1, and so the result follows.

An example of extreme non-classical statistics that violate the Clauser-Horne inequalities consider; $A=S$, $S=B$, but $A=T$ and $T=-B$. The class of all such extreme non-classical statistics is characterized by the violation of

$$c_j(A,S)c_j(S,B) = c_j(A,T)c_j(T,B) \quad \dots (4.8)$$

Again this is a consistency condition for the same reason as for condition (4.3): A as a function of B "via S " should be consistent with, i.e. the same as, A as a function of B "via T ". It is useful to see this in terms of a diagram such as is given in Figure 4.1:

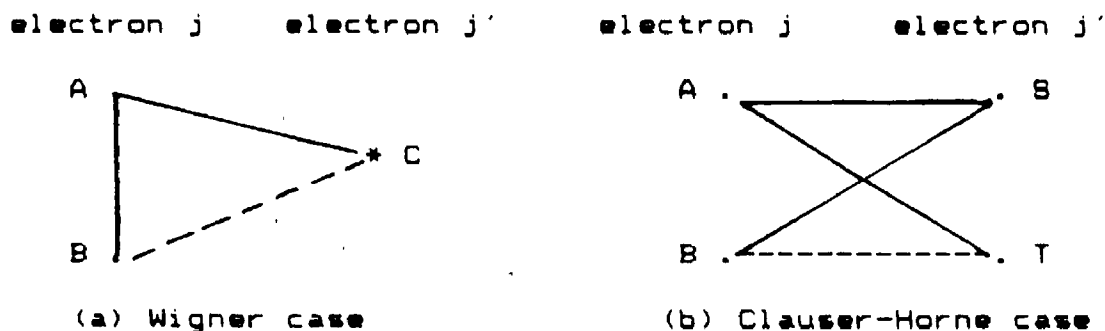


FIGURE 4.1: The solid lines indicate that the corresponding co-equality coefficient is equal to +1, whereas a dotted line indicates a coefficient of -1. When no line directly connects two points in the diagram, then the corresponding coefficient is not given. The consistency constraints are exhausted by the requirement that the product of all the co-equality coefficients along every closed path is +1, or equivalently that they sum to an even integer.

In words, (4.8) expresses the fact that

$$c_j(A,S) = c_j(S,B) \text{ if and only if } c_j(A,T) = c_j(T,B)$$

It is easily checked that (4.8) is therefore equivalent to

$$|c_j(A,S) + c_j(S,B)| + |c_j(A,T) - c_j(T,B)| = 2$$

The derivation of the Clauser-Horne inequalities from here proceeds by taking averages and applying the triangle inequality, to arrive at

$$|C(A,S) + C(S,B)| + |C(A,T) - C(T,B)| \leq 2 \quad \dots (4.9)$$

These inequalities can immediately be recognized as the Clauser-Horne variety because, as an accidental fact, $C(A,S) = \langle A.S \rangle$, etc., (where ' $\langle \rangle$ ' denotes expected value) as is obvious from (4.7).⁷ Therefore (4.9) is recognizable as just the Clauser-Horne inequalities in the form in which Clauser and Horne (1974) originally derived them. To convert (4.9) to the form introduced earlier, simply substitute for the C's using expression (4.7) and

7. This equation is accidental because it depends on the random variables taking on values of -1 or 1. The co-equality coefficients, of course, are independent of such arbitrary calibrations. Of course, the representation of A as a function of B does depend on such assignments and generalizes to $A = c_j(A,B).B + [1-c_j(A,B)]\bar{B}$, where each random variable takes on values v^+ and v^- and $\bar{v} = 1/2(v^+ + v^-)$. But the mutual consistency of these generalized functional relationships leads to the same constraints as before.

simplify. Further generalization of this approach is given in the next section.

The proof of the Clauser-Horne inequalities have two very important advantages over that for the Wigner inequalities: Firstly, there is no need to make use of strict anti-correlation (conservation of spin) to justify the inference that the first electron is spin up in the a -direction from an a -measurement performed on the second of the pair. Secondly, ~~each~~ term of the inequality pertains to information about two compatible quantum mechanical observables.

When conservation of spin is taken into account, the consistency constraints can take a simpler form: For example, take the situation where we consider measurement settings in the a , b , and c directions on both electrons j and j' , as illustrated in Figure 4.2.

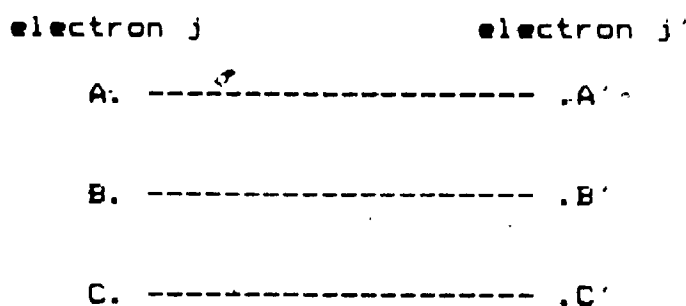


Figure 4.2: The same measurement directions are considered for both electrons.

The dotted lines indicate a strict negative correlation,

i.e. $C(A,A')=-1$, $C(B,B')=-1$, and $C(C,C')=-1$. But this automatically implies that for each j (except on a set of measure zero) $c_j(A,A')=-1$, $c_j(B,B')=-1$, and $c_j(C,C')=-1$. This implies that $C(A,B) = -C(A,B')$, etc., so that the problem as set up for compatible pairs of variables (e.g. A and B') can be translated to a problem concerning random variables applying to one electron only. So the set-up as illustrated in figure 4.2 is really how the 'Wigner' problem should have been set up. In this light, the Clauser-Horne set-up is actually the simplest case in the sense of involving the least number of random variables.

4.3 Farkas' Lemma

In a recent publication, Garg & Mermin (1983) have shown how Bell-type constraints on arbitrary sets of "given" probabilities can be derived from Farkas' Lemma (a species of the so-called double polar theorems) to give necessary and sufficient conditions for the existence of a hidden variable distribution. In this section I will illustrate how Farkas' Lemma can likewise give necessary and sufficient conditions for a given set of co-equality coefficients to yield a hidden variable representation. I will take the simplest case to illustrate the method, the reader can then easily generalize.

Suppose we represent the given co-equality coefficients by a column vector $(c(A,B), c(A,C), c(C,B))^T$.^B The existence of a hidden variable representation reduces to the question whether or not the given vector can be expressed as a convex combination of 'extreme' vectors of the form $(+1, +1, +1)$ subject to the consistency constraints given by (4.3). It is the constraints that make the question non-trivial. If there were no constraints there would

B. The superscript 'T' stands for 'transpose' and converts a row vector into a column vector - it will be omitted where no confusion will result.

always be a hidden variable representation. With the consistency constraints given by (4.3), however, only the extreme vectors $(1,1,1)$, $(1,-1,-1)$, $(-1,1,-1)$, and $(-1,-1,1)$ are allowed, while the other extreme vectors $(-1,-1,-1)$, $(-1,1,1)$, $(1,-1,1)$, and $(1,1,-1)$ are disallowed because they fail the consistency constraint. To ask whether the given vector is a convex combination of the allowable extreme vectors is the same as asking whether there exists a positive solution to the following matrix equation:

$$\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p(\lambda_1) \\ p(\lambda_2) \\ p(\lambda_3) \\ p(\lambda_4) \end{bmatrix} = \begin{bmatrix} C(A,B) \\ C(A,C) \\ C(C,B) \\ 1 \end{bmatrix}$$

Symbolically this equation can be written;

$$\sum_j V_k(\lambda_j) \cdot p(\lambda_j) = C_k \quad \dots (4.10)$$

The 'allowable' extreme vectors discussed above appear as columns in the matrix $V_k(\lambda_j)$, except that the extra element 1 is added to the bottom so that the satisfaction of the equation (4.10) automatically ensures us that $p(\lambda_1) + p(\lambda_2) + p(\lambda_3) + p(\lambda_4) = 1$. All we now need are necessary and sufficient conditions that a positive solution

exists.

FARKAS' LEMMA: There exists a solution $p(\lambda_j)$ that satisfies equation (4.10) such that $p(\lambda_j) \geq 0$, for all j , if, and only if,

$$\sum_k N_k \cdot C_k \geq 0 \quad \dots (4.11)$$

for every row vector N_k such that $\sum_k N_k \cdot \nabla p(\lambda_j) \geq 0$, for all j . (The theorem does not require that $V_k(\lambda_j)$ be a square matrix as is accidentally the case for our example).

Geometrically the theorem has a simple interpretation. Think of the given column matrix C_k on the right hand side of equation (4.10), as well as the columns of the matrix $V_k(\lambda_j)$ as vectors in an abstract vector space. Farkas' lemma then gives necessary and sufficient conditions for the vector C_k to be in the cone generated by the V_k 's. In the 2-dimensional case the content of the lemma is easily visualized. Imagine there are two vectors, a and b , in an abstract vector space, and that we want to find necessary and sufficient conditions for any vector c to be in the cone generated by $\{a, b\}$, call it K . Directly from the definitions we see that c is in K if, and only if, there are $a_1 \geq 0$ and $a_2 \geq 0$ such that $c = a_1 a + a_2 b$. This theorem

9. The cone generated by a set of vectors S is defined as $\text{Cone}(S) = \{x | x = a_1 x_1 + \dots + a_n x_n, \text{ where } x_1, \dots, x_n \text{ are in } S \text{ and } a_1, \dots, a_n \geq 0\}$.

is of little practical use in determining whether c is in K . Farkas' lemma gives a less trivial ~~and~~ more useful result.

Consider the set of all vectors that have a positive dot product with all vectors in K . This set is also a cone and is known as the polar cone of K , and is labelled K° . That is

$$K^\circ = \{x \mid x \cdot y \geq 0, \text{ for all } y \text{ in } K\}$$

A simple illustration of these definitions is seen in the figure below:

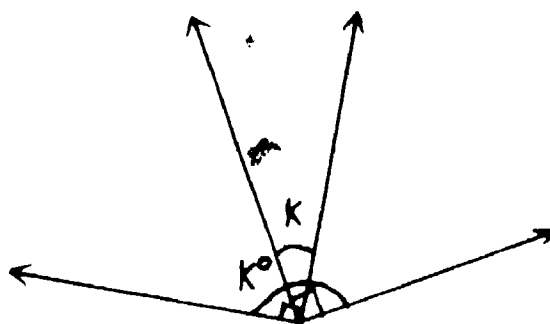


Figure 4.3: An example of a cone K and its polar K° in a 2-dimensional vector space.

In this particular case, it can be observed that not only is K° the polar of K , but conversely K is the polar of K° . It is, in fact, a general truth of finite-dimensional vector spaces that $K^{\circ\circ} = K$, and Farkas' lemma just states that necessary and sufficient conditions for membership of K is simply membership of $K^{\circ\circ}$. It will

now be clear why this is called a 'double polar' theorem.

From the definition above we see that

$$K^{\infty} = \{x \mid x \cdot y \geq 0, \text{ for all } y \text{ s.t. } y \cdot z \geq 0 \text{ for all } z \text{ in } K\}$$

It's a short step from here to see how the statement of Farkas' lemma above follows from the fact that $K^{\infty} = K$.

In the particular example we have been considering, Farkas' lemma produces various permutations of the Wigner inequalities. The corresponding problem, set up for $C(A,S)$, $C(S,B)$, $C(A,T)$, and $C(T,B)$ will yield the Clauser-Horne inequalities, and so on. Many of the inequalities obtained from (4.11) will be trivially satisfied.

The simplest Bell-type problem of all is to ask for the necessary and sufficient conditions for a probability $P(A)$ to be equal to an average over the extreme values 0 and 1. The matrix equation for this problem will be

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} p(\lambda_1) \\ p(\lambda_2) \end{bmatrix} = \begin{bmatrix} P(A) \\ 1 \end{bmatrix}$$

Geometrically, to ask whether there is a positive solution to this equation is to ask whether the given vector lies in the cone generated by the vectors $(1,1)$ and $(0,1)$.

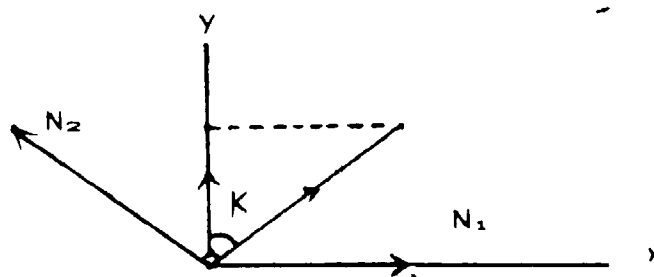


Figure 4.4: An illustration of Farkas' lemma as applied to the problem of determining the conditions under which a single probability can be expressed as an average of extreme values.

Since the second component of the given vector is set at 1, that vector is automatically confined to the horizontal line $y=1$. The necessary and sufficient conditions obtained from Farkas' lemma for the vector $(x,1)$ to lie inside the cone generated by $(1,1)$ and $(0,1)$ is simply that $0 \leq x \leq 1$, i.e. that $0 \leq P(A) \leq 1$. In terms of Farkas' lemma this result, though obvious from the diagram, can be obtained noting that the polar cone is generated by the vectors $N_1 = (1,0)$ and $N_2 = (-1,1)$ and to be inside the original cone requires that $N_1 \cdot (x,1)^T \geq 0$ and $N_2 \cdot (x,1)^T \geq 0$. The first constraint gives $P(A) \geq 0$ while the second reduces to $P(A) \leq 1$.

A less trivial application of the lemma is this: What are necessary and sufficient conditions for the two probabilities $P(A,B)$ and $P(A)$ to be relative frequencies over a common domain? The problem can be formulated thus: We want to know under what conditions there is a positive solution for the unknowns $p(\lambda_1)$, $p(\lambda_2)$, and $p(\lambda_3)$ in the

equation

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p(\lambda_1) \\ p(\lambda_2) \\ p(\lambda_3) \end{bmatrix} = \begin{bmatrix} P(A) \\ P(A.B) \\ 1 \end{bmatrix}$$

Notice that the columns of the matrix are just $(h_1(A), h_1(A.B), 1)$ for the different possible truth valuations associated with the hidden variables λ_1 , λ_2 , and λ_3 . Hence the equation could be written as:

$$\begin{bmatrix} h_1(A) & h_2(A) & h_3(A) \\ h_1(A.B) & h_2(A.B) & h_3(A.B) \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p(\lambda_1) \\ p(\lambda_2) \\ p(\lambda_3) \end{bmatrix} = \begin{bmatrix} P(A) \\ P(A.B) \\ 1 \end{bmatrix}$$

From this way of writing the equation we can think of the question being asked as whether the given vector can be expressed as an average of extreme values. But the important thing to notice is that an affirmative answer to this question requires more than that the components of the given vector be averages of their extreme values, for this simpler condition fails to include the consistency constraints that hold between components.

It is easy to see that in this simple case that the necessary and sufficient conditions for there to exist a positive solution are the inequalities $P(A) \geq 0$, $P(A.B) \geq 0$, $1 - P(A) \geq 0$, $1 - P(A.B) \geq 0$, and $P(A) - P(A.B) \geq 0$ be satisfied. These simple inequalities sometimes appears as axioms of

probability in some axiom schemas - here they are seen as types of Bell inequalities.

It is clear from this example that all the axioms of classical probability theory could be generated from the constraint that the given vector of all probabilities can be expressed as averages over allowed extreme vectors. For classical probability theory, the allowed vectors are those constrained by the consistency requirements of classical logic. (It is through that constraint that the column vector $(0,1,1)$ does not appear in the matrix.) This brings out the intimate connection between classical probability and classical logic.

In terms of the abstract vector space in which this simple problem is formulated, we can draw the cone K generated by the three column vectors in the matrix as is done below:

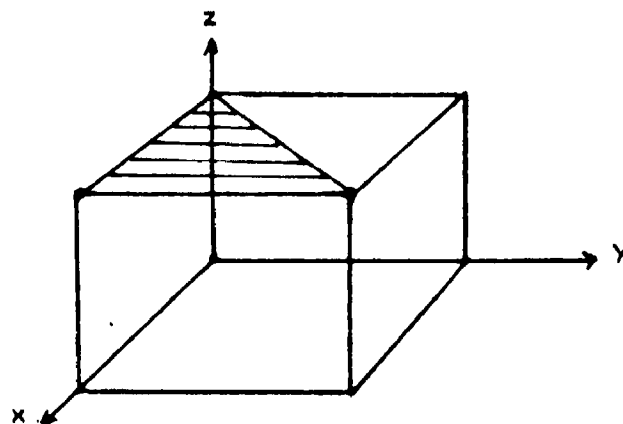


Figure 4.5: A 3-dimensional application of Farkas' lemma.

Because the last element in the given vector is again 1, all such hypothetical vectors are confined to the plane $z=1$. The Bell type inequalities serve to confine the given vector within the simplex of the $z=1$ plane bounded by the (x,y) points $(1,1)$, $(1,0)$ and $(0,0)$.

The difficulty in applying the lemma to more complicated cases is in exhaustively characterizing the class of N_k that give non-trivial constraints. Geometrically it is easy to see that it is sufficient to find the set of normal vectors to each bounding plane of the cone, but in practice this is not so easy.

Secondly, the problem has been moulded to fit the lemma. A more natural statement of the problem would deal only with the abstract vectors without the bottom row of ones. In terms of these restricted vectors the question is simply whether the given vector is in the convex hull generated by the extreme column vectors as they appear in the matrix. That is, whether the given vector can be expressed as a convex combination (a weighted sum where each weight is positive and they all add to one) of the extreme vectors.

We can derive a theorem from Farkas' result to solve this formulation of the problem directly. Take an arbitrary N_k , and indicate all restricted vectors with the bottom '1' removed by the minus as a superscript. Assume

that the last element of this vector is e . Then Farkas' condition is that, for every N_k ,

either $\sum N_k C_k \geq 0$ or $\sum N_k V_k(\lambda_j) < 0$ for some j

But in terms of the restricted vectors, this translates to; For every N_k , and for every real number e ,

either $\sum N_k C_k \geq -e$ or $\sum N_k V_k(\lambda_j) < -e$ for some j

This is because an arbitrary vector N_k can be obtained from its restricted version by adding a number e as the last element. Then by subtracting $1 \cdot e$ from both sides of the inequalities we get the last expression.

The only way this can be violated is if, for some N_k there is some $-e$ such that

$\sum N_k C_k < -e$ or $\sum N_k V_k(\lambda_j) \geq -e$ for all j

But this is the same as

$$\min(\sum N_k V_k(\lambda_j)) > \sum N_k C_k$$

So, finally, the negation of this gives a condition formulated only in terms of the restricted vectors as; For all vectors N_k ,

$$\min(\sum N_k V_k(\lambda_j)) \leq \sum N_k C_k$$

If we write this condition for $-N_k$ we also have that

$$\sum N_k C_k \leq \max(\sum N_k V_k(\lambda_j))$$

This proves the following theorem;

THEOREM: A vector C_k is in the convex hull generated by the vectors $\{V_k(\lambda_j), j=1, \dots, n\}$ if, and only if, for all vectors N_k ,

$$\min(\sum N_k V_k(\lambda_j)) \leq \sum N_k C_k \leq \max(\sum N_k V_k(\lambda_j))$$

This theorem replaces the clumsy application of Farkas' lemma to the problem of determining whether a given vector can be expressed as an average of extreme values under logical constraints. Conceptually the above theorem is neater because it is easy to this of the vector version of the simply condition for scalar averages that they lie between their extreme values. Secondly, it is easy to generate necessary conditions, i.e. Bell-type inequalities, from this theorem as there is no precondition that has to be checked.

4.4 The Classification of Spin-1/2 Correlation Functions

In quantum mechanics, the conditional probabilities of getting spin up in one direction given a certain result in another direction is prescribed as a function of the angle between the two directions. Alternatively, the average

number of times that two results are the same can be given as a function of the angle between the spin directions, at least in the limit as sample numbers become high. When the co-equality coefficient, used as the measure of correlation, is described as a function of angle, I will describe that function as a spin correlation function¹⁰ provided that a few minimal properties are satisfied. The first task is to lay down the properties of spin functions in general and then compare the distinction between the classical and the non-classical cases.

One notion of general spin model can be obtained by generalizing from the following properties of the quantum model. Consider, for instance, the Clauser-Horne situation in which we are interested in the properties of the co-equality coefficient $C(A,S)$ in the limit as $n \rightarrow \infty$. In the quantum mechanical case $C(A,S) \rightarrow -\cos(\hat{a}\hat{s})$. [It is easy to check that this function leads to violations of the Bell inequalities under certain choices of angle.] In the general case we will suppose that $C(A,S) \rightarrow -c(\hat{a}\hat{s})$, where c is a function of angle that has the following properties shared by the quantum mechanical function 'cosine'.

- (1) $c(\theta) = c(-\theta)$, i.e. it doesn't matter whether the angle between two spin directions is
-

10. Accardi (1982) does it differently, I am following Pitowsky (1983b) here.

measured clockwise or anti-clockwise.

(2) Up in the x -direction is the same as down in the $(-x)$ -direction, i.e. $C(X,Y) = -C(-X,Y)$, so that $c(\theta) = -c(\pi-\theta)$, or $c(\theta) + c(\pi-\theta) = 0$.

(3) Conservation of spins: $c(0) = +1$.

(4) It follows from (2) and (3) that $c(\pi) = -1$ and $c(\pi/2) = 0$.

A non-classical spin correlation function is any such function which fails one of the Bell-type inequalities. It will be proven that any function $c(\theta)$ that is continuous at 0 and differentiable in the domain $(0, \pi/2)$ is non-classical if $|c'(\theta)| < 2/\pi$ ($'$ denotes the derivative). The case where $|c'(\theta)| = 2/\pi$ is given by $c(\theta) = 1 - 2|\theta|/\pi$, and is called the classical limit.

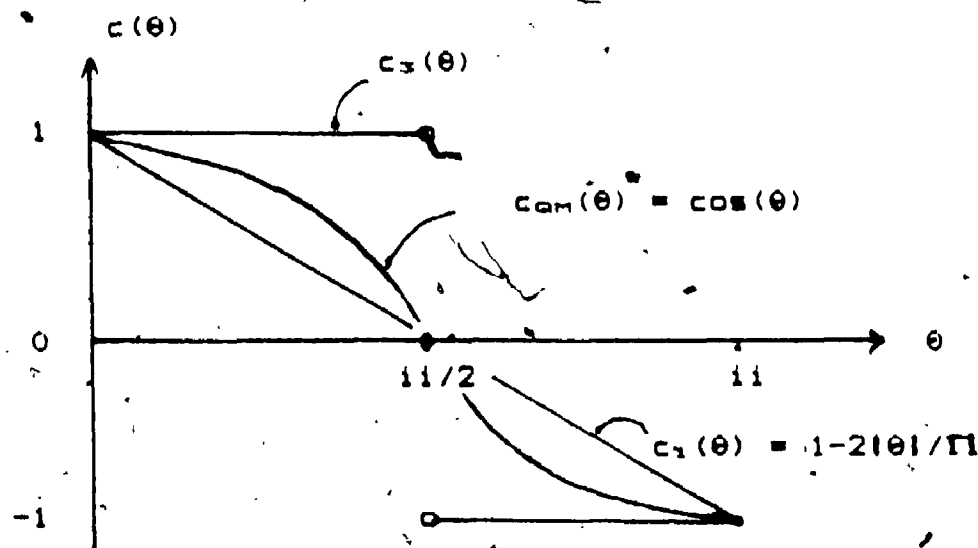


Figure 4.6: Three spin correlation functions; $c_1(\theta)$ is the classical limit, $c_{qm}(\theta)$ is the quantum mechanical case, and $c_3(\theta)$ is the extreme non-classical case in the sense that the cosine function could be "stretched" further from the classical limit to "arrive" at the extreme non-classical case.

We demand of all spin correlation functions that $c(0)=1$ and $-1 \leq c(\theta) \leq 1$. Assuming that $c(\theta)$ is differentiable on the domain $(0, \pi)$ and continuous at $\theta=0$, it follows that $c'(0^+) \leq 0$, where $c'(0^+)$ denotes the derivative of $c(\theta)$ from the right at $\theta=0$. We will prove that if $c(\theta)$ further satisfies the Wigner inequalities, then for all θ in $(0, \pi)$ $|c'(\theta)| \leq -c'(0^+)$, and if $c'(0^+) = -2/\pi$ then $c(\theta) = 1 - 2|\theta|/\pi$.

The function $c(\theta) = 1 - 2|\theta|/\pi$ is then justifiably called the classical limit, because it is the unique classical spin correlation function that is differentiable on $(0, \pi)$ and continuous at 0 and has the maximum slope $-2/\pi$ approaching zero.

The proof of this theorem is as follows: From the Wigner inequalities expressed in the form (4.6) we have that

$$|c(\hat{a}\hat{b}) - c(\hat{b}\hat{c})| \leq 1 - c(\hat{a}\hat{c})$$

Since $\hat{a}\hat{b}$, $\hat{b}\hat{c}$, and $\hat{a}\hat{c}$ are required to be angles between 3 directions in 3-dimensional Euclidean space, they must satisfy the constraints;

$$|\hat{a}\hat{b} - \hat{b}\hat{c}| \leq \hat{a}\hat{c} \leq |\hat{a}\hat{b} + \hat{b}\hat{c}| \text{ and } \hat{a}\hat{b} + \hat{b}\hat{c} + \hat{a}\hat{c} \leq 2\pi$$

Within these bounds we may choose 3 directions a , b , and c

such that $\hat{a}b = \theta+h$, $\hat{b}c = \theta$, and $\hat{a}c = h > 0$, for any θ in the range $(0, \pi)$. Making these substitutions we get;

$$\left| \frac{c(\theta+h) - c(\theta)}{h} \right| \leq \frac{-(c(h) - c(0))}{h}$$

Assuming that the limits $h \rightarrow 0$ exist, we arrive at

$$|c'(\theta)| \leq -c'(0^+).$$

We can further prove that $-c'(0^+) \geq 2/\pi$, but this puts no further constraint on $c'(\theta)$, except that, in the neighborhood of 0, $c(\theta)$ is below the classical limit.

In the limiting case $c'(0^+) = -2/\pi$, we have $|c'(\theta)| \leq 2/\pi$, so it is impossible to go below the classical limit (we have to assume here that $c(\theta)$ is continuous). It is also impossible for $c(\theta)$ to rise above the classical limit in this special case for then it would have to fall again to meet the boundary condition $c(\pi) = -1$, again violating the condition $|c'(\theta)| \leq 2/\pi$.

The class of (well behaved) classical spin correlation functions is limited by the condition $|c'(0^+)| \geq 2/\pi$, the limiting case being unique. This condition alone rules the quantum mechanical case $c_{QM}(\theta) = \cos(\theta)$ as non-classical, of course. All $c(\theta)$, such that $|c'(0^+)| < 2/\pi$ must be non-classical. I.e. all differentiable functions that start above the classical

limit are non-classical. The extreme non-classical case is such that $c'(0^+) = 0$. (See Figure 4.6). Conversely all classical differentiable functions start below the classical limit, but no proof has been given that they must stay below the classical limit [i.e. in the range $(0, \pi/2)$ - but in the range $(\pi/2, \pi)$ 'below' should be replaced by 'above' since for all functions $c(\theta)$, $c(\pi - \theta) = -c(\theta)$.]

d'Espagnat (1979, p.174) suggests that all classical spin correlation functions should be bounded by the classical limit at all points (i.e. bounded above in the domain $(0, \pi/2)$ and below in the domain $(\pi/2, \pi)$). I don't know whether this is true or how one would prove it. However, Pitowsky (1983) has proven that all classical functions are bounded by the classical limit at the discrete points π/n ; $n = 1, 2, \dots$, so there is some hope that such a conjecture is true.

4.5 Consequences for Realism

The general conclusions to be reached from the previous sections are roughly as follows.

- (a) Anyone agreeing that inequality (4.1) is sometimes

violated is faced with some sort of quantum-logical interpretation of quantum mechanics which rejects (4.2), because otherwise we are left with an unpalatable contradiction.

(b) Anti-realists can repudiate the problem by denying that the inequality makes sense in that it presupposes the reality of unmeasured spin values (at most one term in the inequality can be determined experimentally). More will be said about the viability of this option in chapter 5.

(c) Anyone not opting for (a) or (b) must insist that inequality (4.1) is never violated. But there is an argument to the opposite conclusion which he must disarm:

Premise (i). The expectation values of relative frequencies for spin along directions a, b, and c given by quantum theory are such that

$$\langle 1/n \sum_{j \in J} h_j(A.B) \rangle_{\text{qm}} > \langle 1/n \sum_{j \in J} h_j(A.-C) \rangle_{\text{qm}} + \langle 1/n \sum_{j \in J} h_j(B.C) \rangle_{\text{qm}}$$

Premise (ii). The law of large numbers which says that these expectations will be realized (to an arbitrary approximation) at least sometimes, and increasingly often as n becomes larger.

Therefore (iii), there is some set of electrons indexed by J such that inequality (4.1) is violated.

Although we make use of the concept of expectation value (symbolized by $\langle \rangle$), nowhere do the laws governing expectations or probability enter the argument (beyond

those governing relative frequencies). It is unfortunate that the usual formulations of the Bell argument makes use of these laws, since this has lead some to mistakenly believe that postulating non-classical laws of probability is sufficient to resolve the problem. Stairs(1979) rebukes A. Fine for this mistake. In response to challenges to his original claim to extreme realism, classical logic and locality Pitowsky has opted to modify the law of large numbers (1982b). That modification involves the strange thesis that there can be three sets mutually disjoint, but each of probability 1!

Under (c) we are forced to deny one of the premises. One way is to insist that quantum mechanics does not give us expected values simpliciter, but expected values conditional upon measurement. These, after all, are all that are confirmed by experiment. Hence we can agree that quantum mechanics is empirically adequate and still uphold an extreme realist interpretation of quantum mechanics by denying that the hidden variable statistics are the same as the measurement statistics given by quantum theory. This strategy clearly involves clarification of the relationship between hidden variables and measurement outcomes. A detailed analysis must await the next section, for now it suffices to divide premise (1) into two parts:

Premise (i)a: (The experimental facts). If a and b measurements are performed then

$$1/n \sum_{j \in J} h_j(A.B) \rightarrow P_{QM}(A.B), \text{ as } n \rightarrow \infty$$

and if a and c measurements are performed then

$$1/n \sum_{j \in J} h_j(A.-C) \rightarrow P_{QM}(A.-C), \text{ as } n \rightarrow \infty$$

and if b and c measurements are performed then

$$1/n \sum_{j \in J} h_j(B.C) \rightarrow P_{QM}(B.C), \text{ as } n \rightarrow \infty$$

And $P_{QM}(A.B) > P_{QM}(A.-C) + P_{QM}(B.C)$.

Premise (i)b. (Einstein Separability or Conditional Independence.) The antecedents in the above conditionals may be omitted. That is, the hidden unmeasured spin values also conform to the quantum mechanical expectations.

The real import of premise (i)b is that measured statistics can be extrapolated across measurement contexts. More specifically, it says that the quantum mechanical probabilities not only represent the statistics of measurement outcomes, but also the statistics of spin values that are not measured. This premise licenses an extrapolation of statistical regularities, as revealed by the appropriate measurement set-up, to a common probability space. This extrapolation, however, is not possible because it leads to a contradiction. So it follows that quantum statistics are dependent upon the measurement context - 'looking' makes a difference! What this means, in my view, is that the realist cannot succeed in obtaining a cross-situational invariance of the terms appearing in some mathematical model without taking the measurement context into account.

As formulated, Bell's theorem is an argument against an extreme realist but non-quantum-logical interpretation of quantum mechanics. Thus far, there has been no mention of the principle of locality. It will enter the picture as we examine the acceptability of premise (i)b, and so our reconstruction of Bell's argument is not yet complete.

4.6 Enter Locality!

In the last section, premise (i) was separated into two components; (i)a and (i)b. The first simply states the experimental facts and is beyond reproach. A deterministic realist is left only with premise (i)b to attack. But what are the physical ramifications of rejecting premise (i)b (often called conditional independence)? Does the rejection of conditional independence logically obligate the rejection of locality? The answer to this question is a definitive no! Locality is certainly involved, but the crucial question is what assumptions must be added to the principle of locality in order to validate the inference from the experimental facts, (i)a, via the law of large numbers (ii), to the conclusion (iii) - the violation of Bell's inequality. The assumption (conditional independence minus locality) eliminates the so-called

rapport hypotheses, sometimes referred to as the backward light-cone conspiracy theory. This being the case the only viable option open to the realist is to reject the principle of locality. [I am understanding locality here to exclude (cf. Hellman (1982)) the kind of hypothesis held by Costa de Beauregard (1983). Under my usage Costa de Beauregard's theory is non-local.]

To understand the meaning of locality we need to introduce the notion of 'state'.

Definition: For a deterministic hidden variable theory the state history of an electron j is represented by the function $h_j(x,t): S^{(2)} \times R \rightarrow \{0,1\}$, where $S^{(2)}$ is the set of all points on the unit 3-dimensional sphere. The state at a particular time t is the restricted function $h_j(x,t')$.

The usefulness of the notion of 'state' is in making precise some otherwise overly vague use of counterfactual conditionals in the formulation of the Locality condition. The 'state', in other words, is a dispositional property of the system which grounds the truth of or gives substance to the counterfactual conditionals (see section 1.2).

Let " \Rightarrow " be read as "if it were the case that ... then it would be the case that ..." and let $M_j(x,t_i)$ denote the

proposition "a measurement in the x direction was performed on electron j at time t_1 ". The assumption is that a certain restricted class of counterfactual assertions are true or false in virtue of the actual state of the system under discussion.

Deterministic Realism: Counterfactuals of the form

$$M_j(y, t_2) \rightarrow R_j = s_j$$

have a truth value "grounded" in some mind-independent element of reality, or are "objectively" true for some r_j . In particular we assume that, for all directions x and y ,

$$M_j(x, t_1) \rightarrow R_j = r_j \quad \text{iff} \quad h_j(x, t_1) = r_j$$

$$M_j(y, t_2) \rightarrow R_j = s_j \quad \text{iff} \quad h_j(y, t_2) = s_j$$

These counterfactuals are evaluated in a situation that is the same as the actual situation in all respects barring the antecedent itself. In particular, the measurement context for the other electron remains unchanged, as do the states of both electrons. The conditionals are non-backtracking in that it is best to imagine the antecedent as being imposed by some external means completely independent of the system under consideration. In a sense, without some such assumptions the use of counterfactuals would be unhelpful.

For each electron, its state at a particular time determines what the result of some imposed measurement would be for any direction in space. But what determines the state of the electron just before measurement? It is possible that this is determined non-locally. So the assumption that all measurement outcomes are caused locally by the state of the electron does not rule out non-locality in general. It also does not rule out the possibility of a "measurement conspiracy" where the prior state of the system not only determines the outcomes but also the measurement settings in such a way that strange correlations result. In such a case we would have no test of the truth of non-backtracking counterfactual conditionals because no measurement settings would be truly imposed from outside the system.

To formulate a locality condition we need to consider the complete measurement context for both electrons, and how counterfactuals of the form

$$M_1(x, t_1). M_2(y, t_2) \Rightarrow R_1 = r_1 \ \& \ R_2 = s_2$$

are determined by the state of the system. It might be that while the above conditional is true, it is nevertheless false that

$$M_1(z, t_1). M_2(y, t_2) \Rightarrow R_2 = s_2$$

simply because measuring in the z direction instead of the x direction may lead to a change in $h_j(y, t_2)$. The locality condition is designed to eliminate this possibility.

Locality: For all directions x and y ,[†] and for numbers a and b .

$$M_j(x, t_1).M_{j'}(y, t_2) \Rightarrow R_j = a \ \& \ R_{j'} = b$$

$$\text{implies} \quad M_j(z, t_1).M_{j'}(y, t_2) \Rightarrow R_{j'} = b$$

That is, the outcome of any measurement on j' is determined solely by the state of j' , AND that state is not changed by any conditions imposed at j . As a consequence of the locality condition it follows that truth conditions are the same for the class of counterfactual conditionals

$$M_j(x, t_1).M_{j'}(y, t_2) \Rightarrow R_{j'} = b$$

for which direction y is fixed but x can take any value. In particular,

$$[M_j(x, t_1).M_{j'}(y, t_2) \Rightarrow R_{j'} = b] \text{ is true iff } h_{j'}(y, t_2) = b$$

But without the locality condition, we have no right to make such an assumption, even the condition for deterministic realism previously stated would remain true, because there the measurement direction for electron j is identical in all counterfactual situations, viz the actual

one.

The assumption of locality can be seen as doing useful work in placing constraints on the logic of counterfactual conditionals - facts relating to the composite system are expressed in terms of counterfactuals of the type above, and these are related to facts about the component systems, which are also expressed in terms of counterfactuals. Counterfactual conditionals therefore play an important "pre-theoretical" role in helping explore possible cross-situational regularities - in this case among the composite system and its component systems.

As previously mentioned, it is important to interpret these counterfactuals as the non-backtracking variety.¹² Such counterfactuals can be understood as being evaluated in a counterfactual situation in which the antecedent (e.g. $M_i(x).M_j(y)$) is imposed as a "small miracle". Since the discussion has been relativized to a given system, "a small miracle" need not be something supernatural, but simply the imposition of the antecedent condition from outside the system by some means of intervention.¹³ Thus (discounting backward causation for

12. See Lewis (1979)

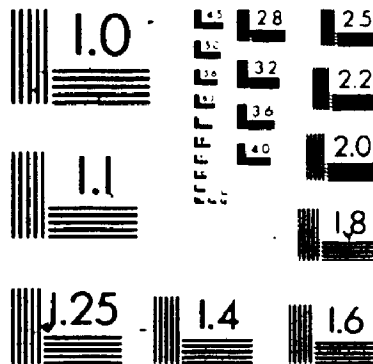
13. See Rescher & Simon (1966) for a theory of causation based on such a notion of intervention.

the moment) nothing prior to the time of measurement is affected by the choice of measurement direction at electron j . It is logically possible, however, that the choice of measurement imposed on electron j at time t_1 does affect the state of electron j' at time t_2 , since time t_2 is later than t_1 . This possibility is ruled out by the principle of locality.

The same assumption also classifies Costa de Beauregard's (1983) theory as non-local. In that theory, the state of electron j' is affected by the imposition of $M_j(z, t_1)$ instead of $M_j(x, t_1)$ (and thereby changing $h_{j'}(y, t_2)$ from what it would have been otherwise) not by any space-like connection between the separated electrons, but by acting along the time-like "zig-zag" via the intersection of their past light-cones. This allows causation to act "bi-directionally" along time-like paths, or as Costa de Beauregard prefers, "causality is arrowless at the micro-level." The above locality assumption, therefore, rests on two separate assumptions: space-like locality (no effect can be received from an event with a space-like separation) and locality from the future (no effects can be received from events in the future light cone).

Even when locality holds, the true statistics over hidden variables may still depend conditionally on what

3 3
OF / DE



measurements are performed (see premise (1)b). It might be that the actual measurement directions are not random, but are 'conspired' to produce non-classical quantum-type statistics, even though the true statistics satisfy the Clauser-Horne inequalities (as they MUST - see section 4.2). In other words, the no-rapport assumption justifies the extrapolation from what occurs naturally within the system to what would happen if measurement settings were imposed upon the system from outside, thereby randomizing any prior "rapport" between the system and the previous measurement settings. This extrapolation is from counterfactual conditionals of the backtracking type (expressing experimental facts) to the non-backtracking type (which express causal propensities of the system).

So the no-rapport assumption states that the measurement statistics are the same as the true underlying statistics of the hidden variables - there is no "rapport" between hidden variables and measurement settings that would make a difference were we able to break that rapport by imposing measurement settings independently. This is not the same as saying that measurement contexts make no difference (Einstein separability), because non-locality can make the measurement context matter even when measurement settings are random. "Measurement conspiracies" and "non-locality" are two subtly different ways in which the measurement context can

count.

But measurement contexts do count if we assume Deterministic Realism, for otherwise we can extapolate across measurement situations and arrive at a contradiction. So one of the assumptions of Locality and No-rapport must be false! Which one?

The rapport hypotheses are less plausible because they imply a strong form of holism. Hidden variables must do two jobs. First of all, they determine the measurement outcomes, but secondly they determine the very direction in which the measurement is made. Non-locality, on the other hand, can produce non-classical results by having (independently determined) measurement directions change the state of the electrons as they pass through the apparatus (measurement disturbance), which in turn affects the state of the other electron non-locally. The measurement directions themselves are completely random, they are determined by causes outside the system. The only non-local action is between electrons. This influence need not act across space-like intervals, as de Beauregard has argued, thereby avoiding the violation of Lorentz invariance.

Locality is sometimes understood as simply no action-at-a-distance across space-like paths [e.g. Hellman, (1982, 1983), Popper (1983)]. Such space-like

non-locality leads to paradoxes and inconsistencies within the special theory of relativity (Savitt (1982) and Savitt & Collier (1983)). However the CPT invariant version of relativity theory does allow a type of relativistic non-locality (Costa de Beauregard, 1983). The present paper purports to show that the question of locality is the correct locus of research. The task for philosophers is to further examine the concept of causality and its applicability to the quantum and relativistic domains. (e.g. Earman, 1983).

The best way to understand some of the points being made is by direct demonstration. In particular, it will be useful here to see how it is possible to produce non-classical statistics within a non-local model. I will take the extreme non-classical case for the Clauser-Horne inequalities illustrated in figure 4.1(b).

Suppose we choose a , b , s , and t to be co-planar directions as illustrated in figure 4.7.

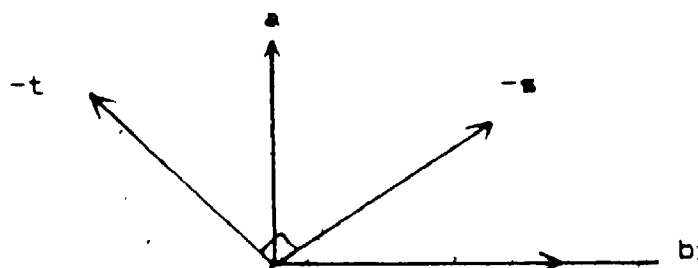


Figure 4.7: One choice of measurement directions for the Clauser-Horne situation that will produce the extreme non-classical

co-equality coefficients as in figure 4.1(b) when these are calculated on the basis of the function $c_3(\theta)$ [see Figure 4.6.].

Experimentally, imagine that we discover that

If $M_j(a).M_j(s)$ then $A = S$

If $M_j(a).M_j(t)$ then $A = T$

If $M_j(b).M_j(s)$ then $B = S$

If $M_j(b).M_j(t)$ then $B = -T$

If these are allowed to hold unconditionally then we easily obtain the contradiction $A=T$ and $A=-T$.

By giving up the locality requirement, it is possible to model non-classical spin systems in accordance with the modelling strategy of path analysis sketched in chapter 3. Let $M(x)$ be a random variable such that $M(x) = 1$ if a spin measurement in the x direction is performed on the electron j , and $M(x) = 0$ otherwise. $M'(x)$ will be the corresponding random variable applied to electron j' . Then we can successfully model this extreme non-classical case as;

$$A = M'(s).S + M'(t).T$$

$$B = M'(s).S - M'(t).T$$

$$S = M(a).A + M(b).B$$

$$T = M(a).A - M(b).B$$

This realist non-quantum-logical model is non-local in

that the basic equations, for A, B, S, and T, must include the random variables $M(a)$, $M(b)$, and $M'(s)$, and $M(t)$, and these random variables represent events occurring at
 14
 space-like separated localities.

This model predicts the correct experimental facts, and these are highly non-classical. But the model also gives the statistics of the underlying hidden variables as classical - as they must be. The point is, then, that non-locality is capable of explaining the experimental manifestation of nonclassical statistics despite the fact that the underlying statistics of the state space is classical.

One objection to deterministic hidden variable models is that they attribute a reality that far transcends the available experimental evidence, and is therefore highly speculative. But the more conservative realist need not tolerate any such surplus "ontological dangles" by instead only accepting a weaker commitment to stochastic hidden variables. Their introduction is all that the evidence can justify, he argues. It is interesting that the realist must still give up locality under this weaker commitment. That topic is treated in the next chapter.

14. To make this explicit, all random variables should be indexed to a particular time or space-time point.

Chapter 5

BELL'S PARADOX AND STOCHASTIC REALISM

5.1 Introduction

Skyrms, in his paper "Counterfactual Definiteness and Local Causation" [Philosophy of Science, 49, 1982], presents a new argument against stochastic local hidden variables in quantum mechanics. He takes a formulation of Locality and Conservation in terms of counterfactual conditionals and concludes on this basis that all stochastic models collapse to being "almost deterministic" in a sense to be explained. From there the usual proofs against deterministic local hidden variables can be applied.

There are other such arguments in the literature [Fine (1982), Suppes & Zanotti (1976), van Fraassen (1983)] - in fact Bell himself (1971) extended his original (1964) proof in just this way. All such proofs, which make extensive use of probability theory, need to assume other conditions - beyond merely Locality and Conservation -

such as to eliminate the possibility of some sort of conspiracy between hidden variables and measurement settings. Commonly this appears as the assumption that hidden variables are stochastically independent of measurement propositions. [E.g. Suppes & Zanotti (1976), van Fraassen (1983).] Usually the meaning of this assumption is not discussed. I will refer to such conditions as "no-rapport" assumptions, for reasons that will be obvious later.

A casual reader, who happens to glance at Skyrms' paper, ought to be quite impressed just from reading the abstract:

Bell's Theorem is proved for locality and conservation formulated in terms of subjunctive conditionals with chance consequents, rather than the usual conditional probability formulation. This brings into sharp focus the minimal counterfactual assumptions needed for Bell's theorem. [My italics]

It appears that, unlike all previous arguments in the literature, Skyrms has proven that no stochastic "rapport" situation can produce quantum-like correlations. And, indeed, he later says explicitly [pp.45-8]

Consider the case in which we have a deterministic hidden variable theory, and the hidden variable is statistically independent of the measurement made. Then it can be shown that the conditional probability $\Pr(R \text{ given } M)$ is equal to the probability of the conditional $\Pr(\text{If } M \text{ then } R)$ It is a virtue of stating the problem in terms of conditionals with chance consequents that the foregoing assumptions are avoided. [my italics]

I gather that the "foregoing assumptions" referred to here are what I will call "deterministic realism" and "no-rapport". Skyrms has successfully avoided the strong assumption of deterministic realism. But he cannot avoid the no-rapport assumption, as I will show later using a simple example.

Skyrms explicitly claims that "No special philosophical theory of subjunctive conditionals is assumed here" [pp.44-45], and in a footnote he adds, "... here I want to use only minimal assumptions". I will show that the meaning of Skyrms' Locality and Conservation conditions are decidedly not neutral with respect to the philosophical theory of subjunctive conditionals. In particular, the meaning changes according to whether the conditionals are taken to be "backtracking" or "non-backtracking". Because non-backtracking conditionals are required for Locality and backtracking is required for Conservation, Skyrms' argument suffers from equivocation. To correct this defect we need to add a "no-rapport" assumption.

But this point does not undercut what I take to be the main aim of Skyrms' paper - to show that only minimal assumptions about the definiteness of counterfactual conditionals are needed for the proof [in response to the contrary charge by Stapp (1971) directed against Bell's

(1971) proof - see Skyrms (1982)1.

5.2 Why Skyrms' Work Is Interesting

I want to say something about why the use of counterfactual conditionals is important and why, therefore, Skyrms' paper is worth discussing in this kind of detail.

Bell (1971) extended his original argument (1964) against deterministic hidden variables to the stochastic case. Both arguments, as Bell formulated them, rely heavily on a classical theory of probability, giving initial plausibility to the view that the paradox might be resolved by developing a non-classical theory of probability. But, for the deterministic case, it can be shown that any non-contextual interpretation of quantum mechanics that embraces the violation of the Bell inequalities, whether through abandoning classical logic or any other reason, has serious difficulties.¹

For suppose we embrace the violation of the Bell-CH

1. Stairs, 1979, appears to have been the first to see this clearly.

2. Fine presents the inequality in slightly different notation and shows how this relates to the original version in Clauser & Horne (1974).

inequality of the form;

$$P(A) + P(S) + P(B.T) - P(A.S) - P(B.S) - P(A.T) > 1$$

where we are now dealing with non-classical probabilities. The propositions A and B refer to spin values in the directions a and b of the first electron, j, and S and T similarly refer to spins in the s and t directions of the second particle, j'. The function h_j is the characteristic function that assigns truth values 0 or 1 to any proposition. If we now assume that these theoretical probabilities connect with relative frequencies in the usual sense that

$$1/n \sum_{j \in J} h_j(A.S) \rightarrow P(A.S) \text{ as } n \rightarrow \infty.$$

then we are in deep trouble. From the mathematical definition of convergence alone, it follows that for any C there is a finite number, N, such that for each term in the above inequality the associated relative frequency stays within C of the quantum mechanical expectation for all $n > N$. If we take C small enough, this ensures us that there is a violation of the Bell-CH inequality in terms of finite relative frequencies. But this is impossible because the Bell-CH inequality is a theorem of classical probability theory and nobody wants to deny that finite relative frequencies satisfy the axioms of classical probability theory! Quantum physicists count up

frequencies the same way as the rest of us. Of course they count up over measurement data, whereas in the context of the Bell inequalities we are counting over unmeasured and measured spin values alike. But assuming that unmeasured values exist (deterministic realism) and measurements do not interfere with the state of the system being measured (the non-contextual interpretation) then there can be no principled difference between measured and unmeasured quantum magnitudes - they are both counted in the same way and, furthermore, measurement statistics give us a reliable indication of the statistics of unmeasured spin values. So if we uphold the connection between probability and relative frequencies as well, in the sense stated above, then appealing to a non-classical theory of probability does us no good.

On the other hand, if it is denied that unmeasured spins, say, have definite values - that the act of measurement brings them into "realization" - then the

3. This is basically the difficulty with the Pitowsky spin model (1983). Within that model, a law of large numbers is proven which says that the relative frequencies will converge to their quantum mechanical values as $n \rightarrow \infty$ "with probability one", where the concept of probability here is the non-standard one developed in the model. But it is clearly a strange sense of probability since each term converges "with probability one" yet we know that at least one term in the inequality must stay well clear of its quantum mechanical value in the limit. It is strange that this term is never the one that is measured! This clearly undercuts the claim that the model is non-contextual.

above problem is repudiated since there are no longer any unmeasured quantities to count over. Such a denial of deterministic realism will be classed as stochastic realism for reasons that will be evident later. It is interesting, therefore, to ask whether there is a proof against stochastic local hidden variable theories that also frees us from the axioms of classical probability in the right way. Bell's proof, or any other proof formulated in terms of probability [e.g. Bell (1971), Fine (1982), Suppes & Zanotti (1976), van Fraassen (1983)], will not do because this still leaves open the possibility that a non-classical theory of probability could unproblematically resolve Bell's paradox. Skyrms' argument can be seen as eliminating this last vestige of hope.

The other important feature of Skyrms' argument is, simply, its use of counterfactual conditionals. This connects Bell's argument with more general discussions of realism in quantum mechanics in a clear and direct way.

Witness Dummett (1976, pp.277-78)

The ... assumption, that, for every test (of some suitably restricted kind), there exists a property which is revealed by the test, and which, at any given time, each object either possesses or fails to possess, is not an operationalist assumption as such, but a realist one. The possession or non-possession of a given property P is what gives substance to the truth of counterfactual statements about what the result of the test T, if it had been applied at a given time, would have been; the bivalence

assumed for statements of the form "The system S has property P" guarantees that one out of every pair of opposite counterfactuals of this kind must be true. Conversely, the supposition that, of each pair of opposite counterfactuals, there must always be one which is true, is tantamount to the realist assumption: the possession of the property can then be equated to the (hypothetical) satisfaction of the test, and will then, in virtue of the supposition about counterfactuals, satisfy the law of bivalence.

With the assumption of realism formulated in this way, it is easy to draw the important distinction between what I will label deterministic realism and stochastic realism.

Let " $O \rightarrow$ " be read as "if it were the case that ... then it would be the case that ..." and let $M_j(x,t)$ denote the proposition "a measurement of spin in direction x was performed on electron j at time t ". Similarly let R refer to the result of a hypothetical spin measurement, R being the case when spin is up and $-R$ corresponding to spin down. $h_j(R) = 1$ says that the result of measurement is spin up for electron j , and $h_j(R) = 0$ says that this is false. Deterministic realism is now the assertion that, for all x in $S^{(2)}$, exactly one of the counterfactual conditionals

$$M_j(x,t) \rightarrow h_j(R) = 1$$

$$M_j(x,t) \rightarrow h_j(R) = 0$$

is true, and the other false. The assumption of realism forces such a law of conditional excluded middle for counterfactual conditionals in that their truth is then

"grounded" in the possession or non-possession of some property.

One way of representing such properties is in terms of a "spin function", $s_j(x,t): S^{(2)} \times \mathbb{R} \rightarrow \{+1/2, -1/2\}$, that assigns a spin value to electron j in every direction in physical space, or equivalently to every point on the unit sphere $S^{(2)}$ in 3-dimensional Euclidean space.⁴ And so we have

Deterministic Realism

$[M_j(x,t) \Rightarrow h_j(R) = 1]$ is true

if, and only if, $s_j(x,t) = +1/2$

Stochastic realism is weaker than deterministic realism in that it only demands that conditionals of the form

$$M_j(x,t) \Rightarrow \langle h_j(R) \rangle = a_j$$

are true for exactly one value of a_j in the closed interval $[0,1]$ (where ' $\langle \rangle$ ' denotes 'expected value'). Because the expected value of a characteristic function is just the probability, such conditionals can be rewritten as

$$M_j(x,t) \Rightarrow P(R) = a_j$$

4. Pitowsky (1983) uses this hidden variable representation.

As Skyrms notes (1982, p46, note 3), this leads to a kind of conditional excluded middle in that exactly one of the counterfactual conditionals

$$M_j(x,t) \Rightarrow P(R) = a_j$$

$$M_j(x,t) \Rightarrow P(R) \neq a_j$$

must be true, for all x in $S^{(2)}$. To represent the properties that "ground" the truth of such counterfactual conditionals involves defining a "state history" and "state" as follows:

Definition: The state history of an electron j is represented as a function

$$p_j(x,t) : S^{(2)} \times \mathbb{R} \rightarrow [0,1].$$

The state at some particular time t is the restricted function

$$p_j(x,t) : S^{(2)} \rightarrow [0,1].$$

For a given time t , the function p_j assigns a propensity value between 0 and 1 to each direction in physical space. It is these propensity properties that now "give substance" to counterfactual conditionals with chance consequents. So now we have that

Stochastic Realism

$$[M_j(x,t) \Rightarrow P(R) = a_j] \text{ is true}$$

if, and only if, $p_j(x;t) = a_j$

Notice that the law of excluded middle no longer applies to conditionals of the form

$$M_j(x,t) \Rightarrow h_j(R) = d$$

where $d = 0$ or 1 . Stochastic realism is weaker than deterministic realism in the sense that spin propositions are not considered to have definite truth values simultaneously for all directions. But stochastic realism still conforms to the same realist motivation (as construed by Dummett) in that the truth of counterfactuals with chance consequences are grounded in the possession or non-possession of a property and conditional excluded middle holds in virtue of this fact.

Within such a framework we can account for the quantum properties of spin measurements made on one electron along the following lines. First suppose that, at any particular time t , each electron has a unique direction of polarization n_t , say, and that the electron state is then completely characterized by the function

$$p_j(x,t) = \cos(\hat{x}\hat{n}_t) \cdot P_j + 1/2[1 - \cos(\hat{x}\hat{n}_t)]$$

where $P_j = 1$ and xn_t is the angle between directions x and n_t . That is $p_j(x) = 1/2 + 1/2\cos(\hat{x}\hat{n}_t) = \cos^2(\hat{x}\hat{n}_t/2)$. We can represent pure states, therefore, as vectors on the

unit sphere in physical space, viz. n_e . n_e is the direction of polarization in the sense that $p_j(x,t) = 1$ if, and only if $x = n_e$.

If $P_j < 1$, then we have a mixed state represented by points inside the unit sphere, of the form $w_1 \cdot n_e + w_2 \cdot -n_e$ where $w_1 = P_j$ and $w_2 = 1 - P_j$. Note that $w_1 \cdot n_e + w_2 \cdot -n_e = P_j \cdot n_e$.

Subjecting electrons in state n_e to an a -measurement will polarize them in the plus or minus a -direction with probabilities $\cos^2(\hat{a}n_e/2)$ and $\sin^2(\hat{a}n_e/2)$ respectively. In this way the stochastic state function $p_j(x)$ determines the state transition probabilities that would occur if an x -measurement were made, as well as the expected outcome of that hypothetical experiment.

5.3 The Reduction of Stochastic Realism to Deterministic Realism

Skyrms (1982) claims that when Locality and Conservation are added to such a picture, the stochastic hidden variables reduce to deterministic ones in the sense that $p_j(x)=0$ or $p_j(x)=1$, for all x in $S^{(2)}$. His argument, as adapted to my notation, goes like this:

I. Locality (L_0):

$$(M_j(x).M_j(y) \Rightarrow F(R'/R) = a)$$

makes it true that $(M_j(x).M_j(y) \Rightarrow P(R') = a)$

which makes it true that $(M_j(y) \Rightarrow F(R') = a)$

The first step in this conditional says that the result of the measurement on electron j does not affect the expected result on j' . The second step in the conditional says that the fact of performing a measurement on j (never mind the result) has no effect on the expected result on j' .⁵

II. Conservation (C_0):

$$(a) \quad (M_j(y).M_{j'}(y) \Rightarrow P(-R) > 0)$$

makes it true that $(M_j(y).M_{j'}(y) \Rightarrow P(R'/-R) = 1)$

$$(b) \quad (M_j(y).M_{j'}(y) \Rightarrow P(R) > 0)$$

makes it true that $(M_j(y).M_{j'}(y) \Rightarrow F(-R'/R) = 1)$

Because at least one of the antecedents of the material conditionals in $C_0(a)$ and $C_0(b)$ must be true, one of the consequents is true. It then follows from L_0 that either

5. Skyrms leaves out the intermediate step, as it plays no part in the proof to follow. I have put it in because it makes it clear how the condition is meant to rule out non-local causation.

$(M_j(y) \rightarrow P(R) = 1)$ or $(M_j(y) \rightarrow P(\neg R) = 1)$ is true. Therefore, from the assumption of counterfactual definiteness (i.e. stochastic realism in this case), the state function is such that $p_j(x) = 1$ or 0 for all directions x in $S^{(2)}$. Hence it appears that, from the assumptions of Locality and Conservation alone, the stochastic case has been reduced to the deterministic case. In actual fact the argument suffers from the fallacy of equivocation. Rather than getting too involved in the formal theory of subjunctive conditionals, I will attempt to prove my point by using some carefully chosen examples.

Skyrms' argument is valid provided the premises are consistently interpreted in terms of counterfactual conditionals that are either backtracking or non-backtracking throughout the argument. This is, I suspect, what Skyrms means when he says "No special philosophical theory of subjunctive conditionals is assumed here". But I will provide an example in which (1) If all conditionals are interpreted as backtracking then L_a is false, despite the fact that Locality clearly holds, and (2) If all conditionals are interpreted as non-backtracking then C_a is false, despite the fact that Conservation does hold for the system.

Non-backtracking counterfactuals can be understood as

being evaluated in a counterfactual situation in which the antecedent (e.g. $M_1(x).M_2(y)$) is imposed as a "small miracle" [see chapter 2]. Since the discussion has been relativized to a given system, "a small miracle" need not be something supernatural, but simply the imposition of the antecedent condition from outside the system by some means of intervention.⁶ Such an evaluation is "non-backtracking" in the sense that the history of the system prior to measurement is not changed.

In contrast, backtracking conditionals describe what happens when the system is left to itself. The hypothetical situation in which it is evaluated is one in which the antecedent of the counterfactual is true, not in virtue of being imposed by any external or divine cause, but because it arose "naturally" in accordance with the internal dynamics of the system. The description of such a situation is obtained by "backtracking" from the antecedent condition, to obtain consistency with the internal laws governing the system. Thus contrued, the distinction I am making is system relative, in that whether or not something is imposed from outside the system is relative to what we take the system to include.

Imagine a set-up where there are two television sets in

6. See Rescher & Simon (1966) for a theory of causation based on such a notion of intervention.

separated locations each of which can be tuned to one of three channels, labelled A, B, and C. A number of occasions, indexed by a set J, are observed and on each occasion each TV flashes either red or green with relative frequency tending to $1/2$. [This example bears some similarity to those described in d'Espagnat (1979) and Mermin (1980).] Whenever both TV's are set to the same channel they flash the same colour. But whenever they are set to different channels, the colours are always different.
7

Suppose that the mechanism that produces these phenomena works as follows: There are in fact only two signal transmitters, ch1 and ch2. It is always TV₁ that receives ch1 and TV₂ that receives ch2. There are two relevant types of situations; (i) when ch1 and ch2 send the same "colour" of signal, and (ii) when they send different "colours". In both cases a pilot signal is sent out; in case (i) it changes the channel settings on both TVs so that they have the same settings and in case (ii) so that they have different settings. Clearly there are no non-local effects here - all correlations are explained in

7. If we wanted to parallel the quantum mechanical case more closely, we would have different colours for same settings - mimicing anti-correlation. But this becomes identical to the example I will use if we place a transducer in one TV that changes "red" to "green" and vice versa.

terms of common causes. And Conservation holds, in that there is perfect correlation of outcomes when "measurement settings" are the same.

(1) Suppose that we interpret all counterfactuals as backtracking. Then Skyrms' Locality condition L_0 must be false. For suppose it is true. I.e suppose

$$(M_1(x).M_2(y) \supset P(R_2/-R_1) = 1),$$

implies $(M_2(y) \supset P(R_2) = 1).$

['R' means that the outcome is red and '-R' means the outcome was green.] Consider case (ii) when ch_1 and ch_2 send different colours - red to ch_1 and green to ch_2 say. Then

$$(M_1(a).M_2(b) \supset P(R_2/-R_1) = 1) \dots (*)$$

is true, and

$$(M_1(b).M_2(b) \supset P(R_2/-R_1) = Q) \dots (**)$$

is also true. But now L_0 leads to a contradiction if we assume that it is true, so it must be false. The statistics governing TV_2 , in other words, are not independent of what the channel dial on TV_1 points to. Skyrms' assumption C_0 is here true, as it should be.

(2) Suppose that all counterfactuals are non-backtracking. This means that they are evaluated in

the hypothetical situation in which the dial settings are imposed from outside the system, irrespective of which settings the pilot signal would have produced. Then (*) and (**) can never be simultaneously true so that no contradiction can be derived by assuming L_0 is true. But now the Conservation assumption is no longer true, for were we to impose the same dial settings in a case where they would normally be different, then we would find different coloured flashes for the same dial settings, contrary to Conservation.

So Skyrms' claim that his argument makes minimal assumptions in being independent of any philosophical theory of subjunctive conditionals is incorrect, as shown by the simple example sketched above. What is needed to eliminate such examples is a "no-rapport" assumption that says that there is no conspiracy between hidden variables and measurement settings. No pilot signals in other words.

The above example was not meant to be a counterexample to Skyrms' argument in which the premisses are true, but the conclusion false. But such a counterexample can be had by adding a "dash of stochasticism" to the model. Assume that in case (ii) - instead of opposite colours - there is some stochastic mechanism that produces an imperfect but high anti-correlation, so that "surface

statistics" remain non-classical. Locality and Conservation continue to hold but nothing forces the model to be deterministic.

5.4 Some Remarks

If we assume locality, conservation and no-rapport the argument goes through, and the stochastic case is reduced to the deterministic case! This correction to Skyrms' proof is minor in the sense that a rapport hypothesis in mechanics is implausible on the grounds that the "pilot signal" would have to act via the "consciousness of the experimenter", since he is necessarily part of the causal chain of events leading to the setting of measurement directions. Nevertheless, such possibilities have been seriously discussed, so in the interest of completeness it is well not to ignore this point.

Take, for instance, the claim made by Bell in his original paper (1964);

Of course, the situation is different if the quantum mechanical predictions are of limited validity. Conceivably they might apply only to experiments in which the settings of the instruments are made sufficiently in advance to allow them to reach some mutual rapport by exchange of signals with velocity less than or equal to that of light.

Recently, experiments by Aspect and co-workers have confirmed that the quantum mechanical predictions still hold for fast rotating polarizers, so that measurement directions are set in space-like separated space-time locations. But this still does not refute the rapport hypothesis; because this does not require faster-than-light signals - all correlations required for the conspiracy can be explained in terms of common cause. In fact it is better explained that way because the conspiracy necessarily requires cooperation with the hidden variables anyway. Conversely, a non-local explanation of the Bell-EPR paradox need not inherit any of the stranger features of the rapport hypothesis such as involving the consciousness of the experimenter. These are common misunderstandings that should be avoided.

And there are always those that come to the realization - through some novel technical apparatus - that Locality, strictly speaking, is not sufficient by itself to derive the Bell inequalities and thereby conclude that Bell's argument does not really impugn Locality after all. But they fail to identify the additional assumptions required to secure the proof and the implausibility of their denials.

As Skyrms emphasizes, the connections between conditional probabilities and counterfactual conditionals

are rather subtle. The condition of Stochastic Realism, as I have formulated it, motivates the introduction of propensities as the properties that ground the truth of subjunctive conditionals with chance consequences. For instance, propensity values assigned to each point on the unit sphere serve to ground the counterfactual expectations for electron spin measurements as they are given by quantum theory itself. This leads to precise constraints on the role and interpretation of theoretical probabilities in quantum mechanics. Conditional propensities can be construed as propensities that would obtain if a certain measurement were to produce a certain result. The non-classical nature of the quantum mechanical spin states is then reflected in the fact that such conditional propensities cannot be embedded into a classical conditional probability function (a Popper function). Popper (1968, appendices iv and v) has proven that such a Popper function generates a Boolean propositional structure under a certain partial ordering. Two questions are hereby raised: (1) Exactly what connections between the logic of counterfactual conditionals and the nature of conditional propensities are implied by the condition of stochastic realism? (2) If measurement always acts as a filter that selects without disturbing the state of the system can it be proven that conditional propensities are always classical

in this sense? [A paper by Sneed (1970) is definitely germane to these questions.]

I suspect that the answer to the second question is yes. The spin model sketched at the end of section 5.2 involves measurement disturbance. This also fits in with the quantum puzzles for correlated electron pairs being seen as an argument that such quantum systems undergo non-local measurement disturbances of some sort. Clearly, both of these questions are worth investigating further.

5.5 A Non-Local Stochastic Model for Electron Spin

As noted earlier, the statistics for the spin of a single electron can be accounted for by assuming that, at any given time, an electron has a unique direction of polarization n_e . This determines a propensity function $p_j(x,t)$ as given by the expression

$$p_j(x,t) = \cos(\hat{x}n_e) + 1/2[1-\cos(\hat{x}n_e)] \quad \dots (5.1)$$

This propensity function gives the probability of state transitions and experimental outcomes. The challenge here is to extend this model to account for the spin statistics of correlated electron pairs in, say, the singlet state.

Let \tilde{n}_1 denote the pure state into which electron₁ collapses after passing through the measurement apparatus at time t_1 , and \tilde{n}_2 similarly for electron₂ at time t_2 . Assume that, at the time of interaction t_0 , the two electrons develop opposite directions of polarization so that they separate with pure states \tilde{n}_0 for electron₁ and $-\tilde{n}_0$ for electron₂. The \tilde{n}_0 will mark the pure state of an electron - represented by a unit vector - as opposed to a mixed state represented by a vector strictly inside the unit sphere.

A model that can account for the known quantum mechanical spin statistics is given follows. Let x_1 be the actual direction of measurement on electron₁, and x_2 for electron₂. (Note that x_1 and x_2 are unit vectors.) Then we can assume that the mixed state of electron₁ after measurement, n_1 , is a function of \tilde{n}_0 and x_1 , and n_2 is a function of $-\tilde{n}_0$ and x_2 , just as for the one-electron case:

$$n_1 = (x_1 \cdot \tilde{n}_0) x_1 \quad \text{and} \quad n_2 = (x_2 \cdot -\tilde{n}_0) x_2 \quad \dots (5.2)$$

Thus, the electron beams directed at each apparatus "collapse" into a mixture determined by the initial state and the direction of measurement, as before.

But we must also assume that n_0 is also a function of the particular outcomes of the measurements, given by the state variables \tilde{n}_1 and \tilde{n}_2 , in order to account for the

strange correlations of quantum mechanics. This is adequately represented by the equation

$$n_0 = w_1(x_1, \tilde{n}_1)x_1 + w_2(x_2, -\tilde{n}_2)x_2 \quad \dots (5.3)$$

where w_1 and w_2 are arbitrary weights strictly between 0 and 1, such that $w_1 + w_2 = 1$.

Now we can prove that this last equation together with the first two reconstructs the quantum statistics. Again we assume the directions of measurement to be x_1 and x_2 , so that there are four possible experimental outcomes in any particular instance, namely $(\tilde{n}_1, \tilde{n}_2) = (\pm x_1, \pm x_2)$. Suppose we take the case in which $\tilde{n}_1 = +x_1$ and ask what the probability is of getting $\tilde{n}_2 = +x_2$ or $\tilde{n}_2 = -x_2$. So we want to solve for n_2 where

$$n_2 = P(\tilde{n}_2 = +x_2 / \tilde{n}_1 = +x_1)x_2 + P(\tilde{n}_2 = -x_2 / \tilde{n}_1 = +x_1)(-x_2) \quad \dots (5.4)$$

However, n_2 is only indirectly a function of \tilde{n}_1 , so we first need maximal information about n_0 , using equation (5.3):

$$\begin{aligned} n_0 &= w_1(x_1, x_1)x_1 + w_2(x_2, -n_2)x_2 \\ &= w_1x_1 + w_2(x_2, -n_2)x_2 \end{aligned}$$

Therefore, by using (5.2),

$$\begin{aligned} n_2 &= (x_2, -n_0)x_2 \\ &= (x_2, -[w_1x_1 + w_2(x_2, -n_2)x_2])x_2 \end{aligned}$$

$$\begin{aligned}
&= \{-w_1(x_1, x_2) + w_2(x_2, n_2)\}x_2 \\
&= -w_1(x_1, x_2)x_2 + w_2(x_2, x_2)\{-w_1(x_1, x_2) + w_2(x_2, n_2)\}x_2 \\
&= -(w_1 + w_1w_2)(x_1, x_2)x_2 + (w_2)^2(x_2, n_2)x_2 \\
&= -(w_1 + w_1w_2 + w_1(w_2)^2)(x_1, x_2)x_2 + (w_2)^3(x_2, n_2)x_2 \\
&= -w_1[1 + w_2 + (w_2)^2 + \dots](x_1, x_2)x_2 \\
&= -[w_1/(1 - w_2)](x_1, x_2)x_2 \\
&= -(x_1, x_2)x_2 \\
&= -\cos(\widehat{x_1 x_2})x_2
\end{aligned}$$

Therefore, if $n_1 = +x_1$,

$$\begin{aligned}
n_2 &= [(1/2 + 1/2\cos(\widehat{x_1 x_2}))x_2 + (1/2 - 1/2\cos(\widehat{x_1 x_2}))(-x_2)] \\
&= \sin^2(\widehat{x_1 x_2}/2)x_2 + \cos^2(\widehat{x_1 x_2}/2)(-x_2)
\end{aligned}$$

Comparing this last expression with (5.4) shows that

$$P(n_2 = +x_2/n_1 = +x_1) = \sin^2(\widehat{x_1 x_2}/2)$$

$$P(n_2 = -x_2/n_1 = +x_1) = \cos^2(\widehat{x_1 x_2}/2)$$

The conditional probability of getting spin up for electron₂ given spin up for electron₁ is the same as the conditional probability of spin down given spin up for sequential measurements on a single electron. In this way the measurement disturbance account of sequential measurements is extended to the two electron case.

B

If we take equations (5.2) and (5.3) as basic, then the

B. See chapter 3 for definitions.

causal action in this model travels two ways along the time-like zig-zag connecting the two measurement events and the common interaction at time t_0 . This is, therefore, a concrete example of the proposal by de Beauregard (1983a, 1983b, 1983c, and 1983d).

It is interesting to note that if this "bi-directional" model is applied religiously to sequential measurements in the directions x_1 at t_1 and x_2 at t_2 as well, we could write:

$$n_2 = (x_2 \cdot \tilde{n}_1) x_2 \quad \dots (5.5)$$

$$n_1 = (x_1 \cdot \tilde{n}_2) x_1 \quad \dots (5.6)$$

If we now take expected values of these two equations, we arrive at

$$n_2 = (x_2 \cdot n_1) x_2$$

$$n_1 = (x_1 \cdot n_2) x_1$$

where

$$n_1 = P(\tilde{n}_1 = +x_1) x_1 + P(\tilde{n}_1 = -x_1) (-x_1)$$

$$n_2 = P(\tilde{n}_2 = +x_2) x_2 + P(\tilde{n}_2 = -x_2) (-x_2)$$

The mutual consistency of equations (5.5) and (5.6) requires that

$$P(\tilde{n}_1 = +x_1) = P(\tilde{n}_1 = -x_1) = 1/2$$

$$P(\tilde{n}_2 = +x_2) = P(\tilde{n}_2 = -x_2) = 1/2$$

This means that the model implies the robustness of absolute probabilities, rather than just the robustness of conditional probabilities normally implied by the non-cyclic models. The usual assumption of an equi-probable distribution is no longer ad hoc. This consequent is, I think, the distinctive feature of such cyclic models, which may have interesting consequences for various issues in the foundations of statistical mechanics. But that is the topic of future study.

5.6 Final Conclusions

The model in the last section conforms to the idea of "arrowless" causation at the microlevel as argued by de Beauregard in that no causal action across space-like paths is assumed. The claim, then, is that such models may be Lorentz invariant. There is, therefore, some promise that such models could lead to further unification of quantum theory with relativity theory. In fact, if Skyrms (1982) is right in claiming that the assumption of stochastic hidden variables (stochastic realism) is already implied by quantum theory itself, then there must be some such model to be had.

Admittedly, the model sketched is highly ad hoc. It is devised only to account for the quantum phenomena within a

narrow domain. The model, by itself, does not fully meet van Fraassen's challenge issued in his (1980) and quoted at the beginning of chapter 1. A convincing realist re-interpretation of the quantum mechanics has not been provided. On the positive side, however, van Fraassen has no convincing counterargument that this strategy cannot be successful, for his construal of the realist position in the guise of Reichenbach's Principle of Common Cause is a straw man.

The corrected view of causation, as developed in chapter 3, shows how Reichenbach's formulation can be weakened so as to apply to any case that can be modelled by a logically consistent set of equations that helps achieve a moderate degree of cross-situational invariance (chapter 1). This is the minimal requirement for the realist to make use of the 'cosmic coincidence' argument to motivate the search for deeper regularities and to project the more robust theoretical functions into new applications. The models sketched in the last section and in chapter 4 prove that the different measurement situations associated with the spin version of the EPR paradox can be described by one unifying system of equations.

The price paid by the realist is that of abandoning the principle of locality. No simple prejudice against non-locality is sufficient to establish this as an

unreasonably high price to pay. But it must be conceded that locality can do some work in unification by establishing the state variables of a composite system as a simple function of the state variables of its component systems. In chapter 4, the locality condition was actually formulated as the requirement that the "test propensities" of the composite system could be reduced to the properties of its components in a simple way. Thus the locality condition certainly has the potential to establish a strong cross-situational invariance of state variables among composite systems and their components. Unfortunately that potential cannot be realized in the quantum mechanical case - that is what the Bell argument proves.

So a cross-situational invariance of theoretical functions must be achieved in some other way. The model presented in the previous section 5.5 is designed to illustrate what such a theory might look like. But the story is far from over. That model must be connected with other such models in diverse domains so that "independent modes of detection" are established for the theoretical functions and this "tightness of fit" is what will count.

If the constructs in some non-local hidden variable spin model could make novel predictions in some diverse domain, then the program would gain some badly needed momentum.

For instance, Tersoff & Bayer, 1983, have shown that the Bose-Einstein statistics can be founded on the classical assumption of "distinguishable particles" if we make some new stochastic assumptions. Kyprianidis et al., 1984, have further argued that these new assumptions can be deduced from the non-local hidden variable theories of Bohm and Vigier. If these developments were to lead to a "tightness of fit" with new or re-interpreted experimental data, then the non-local interpretation of quantum mechanics may become compelling, just as the "crazy" atomic hypothesis finally gained acceptance.

It would then be up to rival theories to surpass these achievements by establishing a greater degree of cross-situational invariance of its theoretical functions.

BIBLIOGRAPHY

- Abrahams J. R. and Coverley, G.P. (1965), Signal Flow Analysis, Pergamon Press.
- Accardi, L. (1982), Foundations of Quantum Probability, Editrice Levrotto & Bella - Torino.
- Accardi, L. & Fedullo, A. (1982), "On the statistical meaning of complex numbers in quantum theory." Lettere al Nuovo Cimento (1982).
- Armstrong, D.M. (1978), Universals and Scientific Realism, 2 vols., Cambridge University Press.
- Armstrong, D.M. (1982), "Laws of Nature as Relations Between Universals and 'As Universals'", in Philosophical Topics, Volume 13, Number 1.
- Ashby, W. Ross (1960), Design for a Brain (2nd edition), John Wiley, N.Y.
- Aspect A., Grangier P., Roger G. (1981), Physical Review Letters 47, p.460.
- Aspect A., Dalibard J., Roger G. (1982), Physical Review Letters 49, p.1804.
- de Beauregard, Costa O. (1983a), "Lorentz and CPT invariances and the EPR correlations" Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.28.
- de Beauregard, Costa O. (1983b), "Einstein-Podolsky-Rosen Correlations: Their Lorentz-and-CPT Invariance", Ms.
- de Beauregard, Costa O. (1983c), "Running backwards the Mermin device: Causality in EPR correlations", American Journal of Physics, 51, 513.
- de Beauregard, Costa O. (1983d), "Lorentz and CPT Invariances and the Einstein-Podolsky-Rosen Correlations", Physical Review Letters, Volume 50, Number 12, 867-869.
- Beckman, P. (1967), Probability in Communication Engineering, Harcourt, Brace & World, Inc.
- Bell, J.S. (1964), "On the Einstein-Podolsky-Rosen paradox" Physics 1 (1964); .pp.195-200.
- Bell, J.S. (1971), "Introduction to the Hidden Variable

- Question", in d'Espagnat (ed.) Foundations of Quantum Mechanics, N. Y.: Academic Press.
- Blalock, H.M., Jr. (1971), Causal Models in the Social Sciences, Adline-Atherton, Chicago.
- Boudon, R. (1974), The Logic of Sociological Explanation, Penguin Books.
- Brush, S.G. (1983), Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager, Princeton University Press.
- Bub, J. (1979), "The Measurement Problem of Quantum Mechanics" in Problems in the Foundations of Physics, G. Toraldo di Francia (ed.) Amsterdam: North-Holland.
- Bunzl, M. (1980), "A note on Doing", Dialogue 19, pp.629-31.
- Bunzl, M. (1984), "A Causal Model for Causal Priority", Erkenntnis 21, pp.31-44.
- Cartwright, N. (1979), "Causal Laws and Effective Strategies", Nous 13, pp.419-37.
- Changeux, J. & Danchin, A. (1976), "Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks", Nature, volume 264, December 23/30.
- Clauser, J.F. & Horne, M.A. (1974), "Experimental Consequences of Objective Local Theories", Physical Review D 10, 526-35.
- Copi, I.M. (1977), "Essence and Accident", in Schwartz (1977), pp.176-191.
- Demopoulos, W. (1982), "The Rejection of Truth-Conditional Semantics by Putnam and Dummett", in Philosophical Topics, Volume 13, Number 1.
- Dretske, F. (1977), "Laws of Nature", Philosophy of Science, 44.
- Dretske, F. (1981), Knowledge and the Flow of Information, MIT Press.
- Dummett M. (1976), "Is Logic Empirical? (1976)", reprinted in Truth and Other Enigmas (1978), Harvard.
- Dwyer, L. (1978), "Time Travel, and Some Alleged Logical Asymmetries Between Past and Future", Canadian Journal

of Philosophy, March 1978.

- Earman, J. (1978), "Fairy Tales vs an Ongoing Story: Ramsey's Neglected Argument for Scientific Realism", Philosophical Studies 33, 195-202.
- Earman, J. (1983), "What is locality? (A skeptical view of some philosophical dogmas)" Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.46.
- Ellis, E. & Sober, E. (1983), "Probabilistic Causality and the Question of Transitivity", Philosophy of Science 50 pp.35-57.
- Eigen, M. & Schuster, P. (1979), The Hypercycle: a Principle of Natural Self-Organisation, Springer-Verlag, N.Y.
- Epstein, W. (ed.) (1977), Stability and Constancy in Visual Perception: Mechanisms and Processes, Wiley Series in Behavior, John Wiley & Sons.
- d'Espagnat, B. (1979), "Quantum Theory and Reality", Scientific American Vol. 241, No.5. pp. 158 - 181.
- Fine, A. (1982), "Hidden Variables, Joint Probability, and the Bell Inequalities." Physical Review Letters Vol.48 No.5 pp.291-5.
- Feller, W. (1968), An Introduction to Probability Theory and its Applications, Vol. 1, John Wiley & Sons.
- Field, H. (1980), "Logic, Meaning, and Conceptual Role", Journal of Philosophy, 74, pp.374-409.
- Forster, M.R. (1978), The Sneedian View of Theories, B.A. Philosophy Honours dissertation, University of Otago.
- Forster, M.R. (1981), The Structure of Biological Science, Technical Report, Number 28(10)·A (1981), Department of Mathematical and Statistical Biology, Czechoslovakian Academy of Sciences, Prague.
- Friedman, M. (1981), "Theoretical Explanation", in Mellor (1981), pp.1-16.
- Friedman, M. (1983), Foundations of Space-Time Theories.
- Garg, A. & Mermin, N.D. (1983), "Farkas's Lemma and the Nature of Reality: Statistical Implications of Quantum Correlations," to appear in Foundations of Physics.

Gibbard, A. and Harper, W. (198), "Counterfactuals and Two Kinds of Expected Utility", in Harper et al. (eds.), Ifs, W.O.S., D. Reidel.

Goodman, N. (1965), Fact, Fiction, and Forecast, 2nd edition, The Bobbs-Merrill Company, Inc.

Gregory, R.L. (1977), Eye and Brain: the Psychology of Seeing, McGraw-Hill, N.Y., Toronto.

Grossberg, S. (1981), "Adaptive Resonance in Development, Percpetion, and Cognition", SIAM-AMS Proceedings, Volume 13, 107-156.

Grossberg, S. (1983), Studies in Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control, Boston Studies in the Philosophy of Science, Volume 70, D. Reidel.

Hacking I. (1982), "Experimentation and Scientific Realism", in Philosophical Topics, Volume 13, Number 1.

Hacking, I. (1983), Representing and Intervening.

Hahlweg, K. (1981), "Progres Through Evolution? An inquiry into the thought of C.H. Waddington", Acta Biotheoretica, 30, pp.103-120.

Hahlweg, K. (1983), The Evolution of Science: a Systems Approach, Ph.D. dissertation, University of Western Ontario.

Harper, W. (1981), "A Sketch of Some Recent Developments in the Theory of Conditionals", in Harper et al. (1981).

Harper, W., Stalnaker, R., & Pearce, G. (1981), Ifs, The University of Western Ontario Series in the Philosophy of Science, Volume 15, D. Reidel.

Harper, W. (1985), "Consilience and Natural Kinds", forthcoming in Philosophy of Science.

Hellman, G. (1982), "Stochastic Einstein-Locality and the Bell Theorems", Synthese, 53, 461-504.

Hellman, G. (1983) "Stochastic Einstein locality and the Bell theorems", Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.88.

Hempel, C.G. (1965), "The Theoreticians' Dilemma", in Aspects of Scientific Explanation, N.Y. Free Press.

- Hoffman, D.D. (1983), "The Interpretation of Visual Illusions", Scientific American, Volume 249, Number 6, pp.154-162.
- Horwich, P. (1982), "Three forms of realism", synthese 51, 181-201.
- Kripke, S. (1977), "Identity and Necessity", in Schwartz (1977), pp.66-101.
- Kuhn, T.S. (1957), The Copernican Revolution, Harvard University Press.
- Kyprianidis, A., Sardelis, D., & Vigier, J.P. (1984), "Causal Non-Local Character of Quantum Statistics", Ms.
- Levins, R. (1968), Evolution in Changing Environments, Princeton University Press.
- Levins, R. (1979), "Coexistence in a Variable Environment", American Naturalist, 114.
- Lewis, D. (1975), "Causation", in Sosa (1975), pp.180-191.
- Lewis, D. (1979), "Counterfactual Dependence and Time's Arrow", Nous.
- Lewis, D. (1983), "New Work for a Theory of Universals", Australasian Journal of Philosophy, Vol.61, No.4, 343-377.
- Lipschutz, S. (1968), Theory and Problems of Linear Algebra (Schaum's Outline Series), McGraw-Hill.
- Lorens, C.S. (1956), "A proof of the non-intersecting loop rule for the solution of linear equations by flow graphs", M.I.T. Quart. Progr. Rep. (Jan. 15).
- McKay, D.M. (1956), "The Epistemological Problem for Automata", in Shannon & McCarthy (1956), pp.235-252.
- McMullen P. & Shephard G.C. (1971), Convex Polytopes and the Upper Bound Conjecture, Cambridge University Press.
- Mellor, D. (1981), Science, Belief and Behaviour, Cambridge University Press.
- Mermin, N.D. (1983), "Pair Distributions and Conditional Independence: Some hints about the structure of strange quantum correlations", Philosophy of Science, 50 pp.359 - 373.

- Mueller, E.W. (1956), "Resolution of the Atomic Structure of a Metal Surface by the Field Ion Microscope", Journal of Applied Physics 27, 474-77.
- Otte, R. (1981), "A Critique of Suppes' Theory of Probabilistic Causality", Synthese 48, 167-189.
- Pavicic, M. (1983), "The Einstein locality without the Bell inequality", Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.161.
- Piaget, J. (1980), Adaptation and Intelligence. Organic Selection and Phenocopy, Translated by S. Eames, University of Chicago Press, Chicago and London.
- Pitowsky, I. (1982a), "Resolution of the EPR and Bell Paradoxes" Physical Review Letters 48, 1299-1302.
- Pitowsky, I. (1982b), "Answers to comments" Physical Review Letters (Oct. 1982).
- Pitowsky, I. (1983a), "Where the Theory of Probability Fails" PSA Vol.2, Proceedings of the Philosophy of Science Association 1982 meeting.
- Pitowsky, I. (1983b), "The Logic of Fundamental Processes: Non-measurable Sets and Quantum Mechanics", Ph.D thesis, university of Western Ontario.
- Pitowsky, I. (1983), Physical Review D, 27, 2316-2326.
- Popper, K. (1968), The Logic of Scientific Discovery, 2nd edition, Hutchinson of London.
- Popper, K. (1983), "A realist view of the EPR experiment", Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.176.
- Prigogine, I. (1980), From Being to Becoming: Time and Complexity in the Physical Sciences, W.H. Freeman, San Francisco.
- Prigogine, I. & Stengers, I. (1984), Order out of Chaos, Bantam Books, N.Y.
- Putnam, H. (1976), "Locke Lecture II" in Meaning and the Moral Sciences, Routledge & Kegan Paul.
- Putnam, H. (1982a), "Why There Isn't a Ready-Made World", Synthese, 51. pp.141-67.
- Putnam, H. (1982b), "Why Reason Can't Be Naturalized",

Synthese, 52. pp.3-23.

Quine, W.V. (1969), "Epistemology Naturalized", reprinted in Ontological Relativity and Other Essays. New York: Columbia University Press.

Quine, W.V. (1974), The Roots of Reference, Open court, La Salle, Illinois.

Ravetz, J. (1966), "The Origins of the Copernican Revolution", Scientific American, October 1966, Volume 215, Number 4, pp.88-103.

Rescher, N. & Simon, H. (1966), "Counterfactuals and Cause" Philosophy of Science (1966), also reprinted in Simon (1977), Models of Discovery D. Reidel.

Robichaud, L.P.A., Boisvert, M., & Robert J. (1962), Signal Flow Graphs and Applications, Prentice-Hall.

Ruse, M. (1973), The Philosophy of Biology, Hutchinson University Library, London.

Ruse, M. (1982), Darwinism Defended: a guide to the evolutionary controversies, Addison-Wesley.

Salmon, W.C. (1970), Statistical Explanation and Statistical Relevance, University of Pittsburgh Press.

Savitt, S. (1982), "Tachyon Signals, Causal Paradoxes, and the Relativity of Simultaneity" in T. Nickles (ed.): PSA, 1982, vol.1 East Lansing Michigan: Philosophy of Science Association.

Savitt, S.F. & Collier, J. (1983), "Tachyons and causal theories of space-time" Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.210.

Scheibe, E. (1983), "What kind of hidden variables are excluded by Bell's inequality?", Abstracts of the 7th International Congress of Logic, Methodology, and Philosophy of Science, Vol.8, p.213.

Schwartz, S.P. (ed.) (1977), Naming, Necessity, and Natural Kinds, Cornell University Press.

Segal, M.H., Campbell, D.T., & Herskovits, M.J. (1966), The Influence of Culture on Visual Perception, The Bobbs-Merrill Company.

Segal, M.H., Campbell, D.T., & Herskovits, M.J. (1969), "Cultural Differences in the Perception of Geometric

- Illusions", in Tibbetts, P. (ed.), Perception: Selected Readings in Science and Phenomenology, Quadrangle/ The New York Times Book Co, N.Y.
- Shannon, C.E. & McCarthy, J. (eds) (1956), Automata Studies, Annals of Mathematical Studies, Number 34, Princeton University Press.
- Shimony, A. & Hultgren, B.O. (1977), "The lattice of verifiable propositions of the spin-1 system", Journal of Mathematical Physics, volume 18, Number 3, March 1977.
- Shrader, D.W. (1977), "Causation, Explanation, and Statistical Relevance", Philosophy of Science 44, pp.136-145.
- Sklar, L. (1982), "Saving the Noumena", Philosophical Topics, Volume 13, Number 1.
- Skyrms, B. (1977), "Resiliency, Propensities, and Causal Necessity", Journal of Philosophy, LXXIV, 704-713.
- Skyrms, B. (1980), Causal Necessity, Yale University Press: New Haven.
- Skyrms, B. (1982), "Counterfactual Definiteness and Local Causation", Philosophy of Science 49, pp.43-50.
- Skyrms, B. (1984), EPR: Lessons for Metaphysics, forthcoming in Midwest Studies in Philosophy.
- Sneed, J.D. (1970), "Quantum mechanics and classical probability theory", Synthese 21, 34-64.
- Sneed, J.D. (1971), The Logical Structure of Mathematical Physics, D. Reidel.
- Sober, E. & Lewontin, R. (1982), "Artifact, Cause and Genic Selection", Philosophy of Science, June 1982.
- Sosa, E. (ed.) (1975), Causation and Conditionals, Oxford University Press.
- Stairs, A. (1979), "On Arthur Fine's Interpretation of Quantum Mechanics" Synthese 42 pp.91-100.
- Stabler, E. (1982), "Naturalized Epistemology and Metaphysical Realism: A Response to Rorty and Putnam", in Philosophical Topics, Volume 13, Number 1.
- Stapp, H.P. (1971), "S-matrix Interpretation of Quantum Theory", Physical Review D 3: 1303-20.

- Suppe, F. (1977), The Structure of Scientific Theories, 2nd edition, University of Illinois Press.
- Suppes, P. (1970), A Probabilistic Theory of Causality, Amsterdam: North Holland Publishing Company.
- Suppes & Zanotti (1976), "On the determinism of Hidden Variable Theories with Strict Correlation and Conditional Statistical Independence of Observables" in Suppes (ed.) Logic and Probability in Quantum Mechanics, Dordrecht: D. Reidel.
- Tersoff, J. & Bayer, D. (1983), "Quantum for Statistics for Distinguishable Particles", Physical Review Letters, Volume 50, Number 8, pp.553-4.
- Tichý, P. (1978), "A New Theory of Subjunctive Conditionals", Synthese, 37, 433-457.
- Tooley, M. (1977), "The Nature of Laws", Canadian Journal of Philosophy, 7.
- Tweedale, M. (1982), "Universals and Laws of Nature", in Philosophical Topics, Volume 13, Number 1.
- van Fraassen, B.C. (1980), The Scientific Image, Clarendon Press, Oxford.
- van Fraassen, B.C. (1983), "The Charybdis of Realism: Epistemological Foundations of Bell's Inequality Argument", Synthese.
- Wiberg, D.M. (1971), Theory and Problems of State Space and Linear Systems (Schaum's Outline Series), McGraw-Hill.
- Wimsatt, W. (1980), "Randomness and Perceived Randomness in Evolutionary Biology", Synthese, 43, 287-239.
- Wright, von G.H. (1975), "On the Logic and Epistemology of the Causal Relation", in Sosa (1975), pp.95-113.
- Wright, S. (1968), "Path Analysis: Theory", Chapter 13 in Evolution and the Genetics of Populations, Volume 1; Genetic and Biometric Foundations, Chicago University Press.

END

27103185

FIN